

This is a repository copy of *Developing and evaluating packages to support implementation of quality indicators in general practice : the ASPIRE research programme, including two cluster RCTs*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/159886/>

Version: Published Version

Article:

Foy, Robbie, Willis, Tom, Glidewell, Liz orcid.org/0000-0003-2519-2654 et al. (28 more authors) (2020) Developing and evaluating packages to support implementation of quality indicators in general practice : the ASPIRE research programme, including two cluster RCTs. Programme Grants for Applied Research. ISSN 2050-4322

<https://doi.org/10.3310/pgfar08040>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

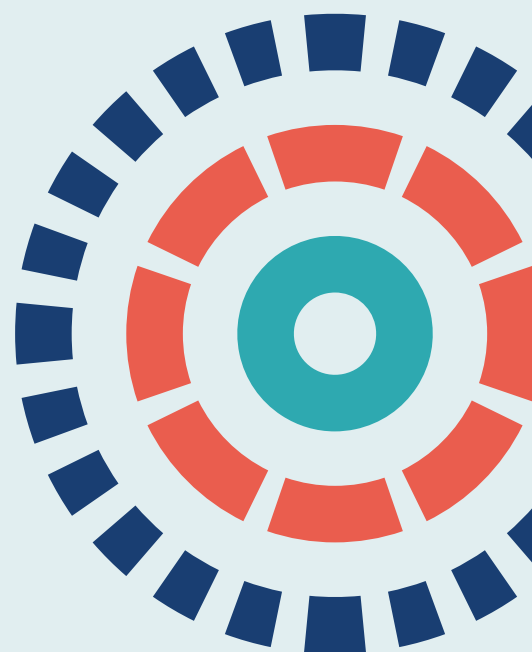
Programme Grants for Applied Research

Volume 8 • Issue 4 • March 2020

ISSN 2050-4322

Developing and evaluating packages to support implementation of quality indicators in general practice: the ASPIRE research programme, including two cluster RCTs

*Robbie Foy, Thomas Willis, Liz Glidewell, Rosie McEachan, Rebecca Lawton,
David Meads, Michelle Collinson, Cheryl Hunter, Claire Hulme, Robert West,
Vicky Ward, Suzanne Hartley, Paul Carder, Sarah Alderson, Michael Holland,
Peter Heudtlass, Daniele Bregantini, Laetitia Schmitt, Susan Clamp, Tim Stokes,
Emma Ingleson, Martin Rathfelder, Stella Johnson, Judith Richardson,
Bruno Rushforth, Duncan Petty, Armando Vargas-Palacios, Gemma Louch,
Jane Heyhoe, Ian Watt and Amanda Farrin*



Developing and evaluating packages to support implementation of quality indicators in general practice: the ASPIRE research programme, including two cluster RCTs

Robbie Foy^{1*}, Thomas Willis¹, Liz Glidewell¹,
Rosie McEachan², Rebecca Lawton^{2,3}, David Meads⁴,
Michelle Collinson⁵, Cheryl Hunter⁶, Claire Hulme⁴,
Robert West⁷, Vicky Ward¹, Suzanne Hartley⁵,
Paul Carder⁸, Sarah Alderson¹, Michael Holland⁵,
Peter Heudtlass⁹, Daniele Bregantini¹⁰,
Laetitia Schmitt¹¹, Susan Clamp¹², Tim Stokes¹³,
Emma Ingleson⁵, Martin Rathfelder¹⁴,
Stella Johnson⁸, Judith Richardson¹⁵,
Bruno Rushforth¹⁶, Duncan Petty¹⁷,
Armando Vargas-Palacios⁴, Gemma Louch¹⁸,
Jane Heyhoe², Ian Watt¹⁹ and Amanda Farrin⁵

¹Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

²Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

³Department of Psychology, University of Leeds, Leeds, UK

⁴Academic Unit of Health Economics, University of Leeds, Leeds, UK

⁵Clinical Trials Research Unit, Leeds Institute for Clinical Trials Research, University of Leeds, Leeds, UK

⁶University of East London, London, UK

⁷Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK

⁸NHS Bradford Districts Clinical Commissioning Group, Bradford, UK

⁹Centre for Health Research & Evaluation, National Pharmacy Association, Lisbon, Portugal

¹⁰Management School, University of Liverpool, Liverpool, UK

¹¹Centre for Health Economics, University of York, York, UK

¹²Yorkshire Centre for Health Informatics, University of Leeds, Leeds, UK

¹³Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

¹⁴Socialist Health Association, Little Sutton, Ellesmere Port, UK

¹⁵National Institute for Health and Care Excellence, London, UK

¹⁶Foundry Lane Surgery, Leeds, UK

¹⁷Faculty of Life Sciences, University of Bradford, Bradford, UK

¹⁸Yorkshire Quality and Safety Research Group, Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK

¹⁹Department of Health Sciences, Hull York Medical School, University of York, York, UK

*Corresponding author

Declared competing interests of authors: As chief investigator, Robbie Foy was responsible for commissioning and funding Prescribing Support Services Ltd (PSS) (Shipley, UK) to conduct the educational outreach visits as part of the implementation package. PSS (which is Duncan Petty's practice pharmacy company) was paid to provide pharmacists to run the educational outreach visits. The commissioning of PSS followed established NHS commissioning guidelines. Robbie Foy also reports that he is a member of the Dissemination Centre Advisory Group (2015–present). David Meads reports that he was a member of the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) European Economic and Social Committee EESC Methods Group (2014–17) and NIHR HTA EESC Panel (2013–17). Amanda Farrin reports that she was a member of the HTA Antimicrobial Resistance Themed Call Board (2013–14), HTA Clinical Trials Board (2014–18), HTA Efficient Study Designs Board (2014), HTA Flu Themed Call Board (2009–11), HTA Funding Board Policy Group (formerly Commissioning Strategy Group) (2014–18), HTA Obesity Themed Call Board (2010–10), HTA Pandemic Influenza Board (2011–11), HTA Primary Care Themed Call Board (2013–14), HTA Surgery Themed Call Board (2012–13), HTA Trauma Themed Call Board (2007–8) and Rapid Trials and Add on Studies Board (2012). Claire Hulme was a member of the NIHR HTA Commissioning Board (2013–17). Bruno Rushforth was in receipt of NIHR In-Practice Fellowship Funding during the conduct of the programme. Robert West reports membership of the Health Services and Delivery Research Researcher Led Panel (2017–20) and the Public Health Research Funding Board (2011–17). Michelle Collinson reports that since September 2018 she has been a member of the Research for Patient Benefit regional funding panel for Yorkshire and the North East.

Published March 2020

DOI: 10.3310/pgfar08040

This report should be referenced as follows:

Foy R, Willis T, Glidewell L, McEachan R, Lawton R, Meads D, *et al.* Developing and evaluating packages to support implementation of quality indicators in general practice: the ASPIRE research programme, including two cluster RCTs. *Programme Grants Appl Res* 2020;**8**(4).

Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full PGfAR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/pgfar. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Programme Grants for Applied Research* journal

Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was established in 2006 to fund collaborative, multidisciplinary programmes of applied research to solve health and social care challenges. Findings are expected to provide evidence that lead to clear and identifiable patient benefits, in the relatively near future.

PGfAR is researcher led and does not specify topics for research; however, the research must be in an area of priority or need for the NHS and the social care sector of the Department of Health and Social Care, with particular emphasis on health and social care areas that cause significant burden, where other research funders may not be focused, or where insufficient funding is available.

The programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director. For more information about the PGfAR programme please visit the website: <https://www.nihr.ac.uk/explore-nihr/funding-programmes/programme-grants-for-applied-research.htm>

This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-1209-10040. The contractual start date was in December 2014. The final report began editorial review in February 2018 and was accepted for publication in November 2019. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, CCF, NETSCC, PGfAR or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the PGfAR programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2020. This work was produced by Foy *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Programme Grants for Applied Research* and NIHR Journals Library

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Senior Clinical Researcher, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Developing and evaluating packages to support implementation of quality indicators in general practice: the ASPIRE research programme, including two cluster RCTs

Robbie Foy^{1*}, Thomas Willis¹, Liz Glidewell¹, Rosie McEachan², Rebecca Lawton^{2,3}, David Meads⁴, Michelle Collinson⁵, Cheryl Hunter⁶, Claire Hulme⁴, Robert West⁷, Vicky Ward¹, Suzanne Hartley⁵, Paul Carder⁸, Sarah Alderson¹, Michael Holland⁵, Peter Heudtlass⁹, Daniele Bregantini¹⁰, Laetitia Schmitt¹¹, Susan Clamp¹², Tim Stokes¹³, Emma Ingleson⁵, Martin Rathfelder¹⁴, Stella Johnson⁸, Judith Richardson¹⁵, Bruno Rushforth¹⁶, Duncan Petty¹⁷, Armando Vargas-Palacios⁴, Gemma Louch¹⁸, Jane Heyhoe², Ian Watt¹⁹ and Amanda Farrin⁵

¹Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

²Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

³Department of Psychology, University of Leeds, Leeds, UK

⁴Academic Unit of Health Economics, University of Leeds, Leeds, UK

⁵Clinical Trials Research Unit, Leeds Institute for Clinical Trials Research, University of Leeds, Leeds, UK

⁶University of East London, London, UK

⁷Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK

⁸NHS Bradford Districts Clinical Commissioning Group, Bradford, UK

⁹Centre for Health Research & Evaluation, National Pharmacy Association, Lisbon, Portugal

¹⁰Management School, University of Liverpool, Liverpool, UK

¹¹Centre for Health Economics, University of York, York, UK

¹²Yorkshire Centre for Health Informatics, University of Leeds, Leeds, UK

¹³Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

¹⁴Socialist Health Association, Little Sutton, Ellesmere Port, UK

¹⁵National Institute for Health and Care Excellence, London, UK

¹⁶Foundry Lane Surgery, Leeds, UK

¹⁷Faculty of Life Sciences, University of Bradford, Bradford, UK

¹⁸Yorkshire Quality and Safety Research Group, Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK

¹⁹Department of Health Sciences, Hull York Medical School, University of York, York, UK

*Corresponding author r.foy@leeds.ac.uk

Background: Dissemination of clinical guidelines is necessary but seldom sufficient by itself to ensure the reliable uptake of evidence-based practice. There are further challenges in implementing multiple clinical guidelines and clinical practice recommendations in the pressurised environment of general practice.

Objectives: We aimed to develop and evaluate an implementation package that could be adapted to support the uptake of a range of clinical guideline recommendations and be sustainably integrated within general practice systems and resources. Over five linked work packages, we developed 'high-impact' quality indicators to show where a measurable change in clinical practice can improve patient outcomes (work package 1), analysed adherence to selected indicators (work package 2), developed an adaptable implementation package (work package 3), evaluated the effects and cost-effectiveness of adapted implementation packages targeting four indicators (work package 4) and examined intervention fidelity and mechanisms of action (work package 5).

Setting and participants: Health-care professionals and patients from general practices in West Yorkshire, UK.

Design: We reviewed recommendations from existing National Institute for Health and Care Excellence clinical guidance and used a multistage consensus process, including 11 professionals and patients, to derive a set of 'high-impact' evidence-based indicators that could be measured using routinely collected data (work package 1). In 89 general practices that shared data, we found marked variations and scope for improvement in adherence to several indicators (work package 2). Interviews with 60 general practitioners, practice nurses and practice managers explored perceived determinants of adherence to selected indicators and suggested the feasibility of adapting an implementation package to target different indicators (work package 3). We worked with professional and patient panels to develop four adapted implementation packages. These targeted risky prescribing involving non-steroidal anti-inflammatory and antiplatelet drugs, type 2 diabetes control, blood pressure control and anticoagulation for atrial fibrillation. The implementation packages embedded behaviour change techniques within audit and feedback, educational outreach and (for risky prescribing) computerised prompts. We randomised 178 practices to implementation packages targeting either diabetes control or risky prescribing (trial 1), or blood pressure control or anticoagulation (trial 2), or to a further control (non-intervention) group, and undertook economic modelling (work package 4). In trials 1 and 2, practices randomised to the implementation package for one indicator acted as control practices for the other package, and vice versa. A parallel process evaluation included a further eight practices (work package 5).

Main outcome measures: Trial primary end points at 11 months comprised achievement of all recommended levels of glycated haemoglobin, blood pressure and cholesterol; risky prescribing levels; achievement of recommended blood pressure; and anticoagulation prescribing.

Results: We recruited 178 (73%) out of 243 eligible general practices. We randomised 80 practices to trial 1 (40 per arm) and 64 to trial 2 (32 per arm), with 34 non-intervention controls. The risky prescribing implementation package reduced risky prescribing (odds ratio 0.82, 97.5% confidence interval 0.67 to 0.99; $p = 0.017$) with an incremental cost-effectiveness ratio of £2337 per quality-adjusted life-year. The other three packages had no effect on primary end points. The process evaluation suggested that trial outcomes were influenced by losses in fidelity throughout intervention delivery and enactment, and by the nature of the targeted clinical and patient behaviours.

Limitations: Our programme was conducted in one geographical area; however, practice and patient population characteristics are otherwise likely to be sufficiently diverse and typical to enhance generalisability to the UK. We used an 'opt-out' approach to recruit general practices to the randomised trials. Subsequently, our trial practices may have engaged with the implementation package less than if they had actively volunteered. However, this approach increases confidence in the wider applicability of trial findings as it replicates guideline implementation activities under standard conditions.

Conclusions: This pragmatic, rigorous evaluation indicates the value of an implementation package targeting risky prescribing. In broad terms, an adapted 'one-size-fits-all' approach did not consistently work, with no improvement for other targeted indicators.

Future work: There are challenges in designing 'one-size-fits-all' implementation strategies that are sufficiently robust to bring about change in the face of difficult clinical contexts and fidelity losses. We recommend maximising feasibility and 'stress testing' prior to rolling out interventions within a definitive evaluation. Our programme has led on to other work, adapting audit and feedback for other priorities and evaluating different ways of delivering feedback to improve patient care.

Trial registration: Current Controlled Trials ISRCTN91989345.

Funding: This project was funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research programme and will be published in full in *Programme Grants for Applied Research*; Vol. 8, No. 4. See the NIHR Journals Library website for further project information.

Contents

List of tables	xiii
List of figures	xv
List of boxes	xvii
List of supplementary material	xix
List of abbreviations	xxi
Plain English summary	xxiii
Scientific summary	xxv
SYNOPSIS	1
Work package 1: developing 'high-impact' quality indicators for primary care	1
<i>Background</i>	1
<i>Methods and results</i>	2
<i>Discussion</i>	4
<i>Conclusion</i>	5
Work package 2: variations in achievement of evidence-based, high-impact quality indicators in general practice	5
<i>Background</i>	5
<i>Methods</i>	6
<i>Results</i>	6
<i>Discussion</i>	8
Work package 3a: using the theoretical domains framework to understand adherence to multiple quality indicators in primary care	9
<i>Background</i>	9
<i>Methods</i>	10
<i>Results</i>	11
<i>Discussion</i>	13
<i>Conclusion</i>	14
Work package 3b: developing an adaptable implementation package for indicators in primary care	14
<i>Background</i>	14
<i>Methods</i>	15
<i>Results</i>	16
<i>Discussion</i>	19
<i>Conclusion</i>	19
Work package 4a: a cluster-randomised evaluation of an adaptable implementation package targeting 'high-impact' evidence-based indicators in primary care	24
<i>Background</i>	24
<i>Methods</i>	24
<i>Results</i>	29
<i>Discussion</i>	32
<i>Conclusion</i>	34

CONTENTS

Work package 4b: economic evaluation	34
<i>Introduction</i>	34
<i>Methods</i>	34
<i>Results</i>	39
<i>Discussion</i>	46
<i>Conclusion</i>	47
Work package 5: a multimethod process evaluation of an adaptable implementation strategy	47
<i>Background</i>	47
<i>Methods</i>	47
<i>Results</i>	50
<i>Discussion</i>	54
<i>Conclusion</i>	55
Patient and public involvement	55
<i>Roles and experience within ASPIRE</i>	55
<i>The role of patient and public involvement in implementation research</i>	56
Conclusions	59
<i>Implications for practice</i>	60
<i>Implications for future research</i>	60
Acknowledgements	63
References	67
Appendix 1 Supplementary tables and figures	75

List of tables

TABLE 1 Quality indicators explored in qualitative interviews with primary care staff (WP3a)	10
TABLE 2 Allocation of interview topics by primary care staff (WP3a)	12
TABLE 3 Theoretical determinants of adherence to QIs and their prioritisation for intervention development (WP3b)	17
TABLE 4 Candidate BCTs mapped to determinants of adherence ordered by core, prominent and less evident determinants identified during interview study (WP3b)	18
TABLE 5 Full description of intervention content by delivery mechanism and QI	20
TABLE 6 Overview of cluster-randomised evaluation (WP4a): trial 1	24
TABLE 7 Overview of cluster-randomised evaluation (WP4a): trial 2	24
TABLE 8 Primary and secondary outcomes in cluster-randomised evaluation of implementation package (WP4a)	26
TABLE 9 Key sample size assumptions for cluster-randomised evaluation of implementation package (WP4a)	27
TABLE 10 Primary outcome achievement: baseline rates, adjusted outcome rates and adjusted odds (WP4a)	31
TABLE 11 Economic evaluation: fixed costs of ASPIRE implementation package (WP4b)	35
TABLE 12 Economic evaluation: mean outcomes for the composite risky prescribing indicator (WP4b)	39
TABLE 13 Economic evaluation: mean PSA outcomes at the practice level for the composite BP control indicator (WP4b)	43
TABLE 14 Economic evaluation: mean individual-level summary results from the UKPDS model (WP4b)	45
TABLE 15 Summary of fidelity data collection methods (WP5)	48
TABLE 16 Primary outcome achievement rate by uptake of outreach visit, with non-intervention group for comparison (WP5)	53
TABLE 17 Description of developed implementation package following the template for intervention description and replication (TiDieR) checklist (WP3b)	75
TABLE 18 Behaviour change techniques excluded from intervention development or intended but not subsequently operationalised (WP3b)	78

TABLE 19 Cluster-randomised evaluation: characteristics of the non-randomised and randomised practices (WP4a)	80
TABLE 20 Cluster-randomised evaluation: general practice characteristics at baseline by trial arm (WP4a)	82
TABLE 21 Cluster-randomised evaluation: patient characteristics at baseline by trial arm (all patients) (WP4a)	83
TABLE 22 Primary outcome analysis: diabetes control (WP4a)	84
TABLE 23 Primary outcome analysis: risky prescribing (WP4a)	85
TABLE 24 Primary outcome analysis: BP control (WP4a)	87
TABLE 25 Primary outcome analysis: anticoagulation in AF (WP4a)	88
TABLE 26 Number of practices in each trial arm where at least one outreach visit took place (WP5)	89
TABLE 27 Number of outreach visits held at each practice (by trial arm) (WP5)	89
TABLE 28 Intervention fidelity level of practices by trial arm (WP4a)	89
TABLE 29 Summary of delivery and receipt of intervention (WP4a)	89
TABLE 30 Perceptions of implementation package components (practices anonymised and assigned pseudonyms) (WP5)	91
TABLE 31 Reported actions in response to implementation package components (WP5)	93

List of figures

FIGURE a Research pathway diagram	xxvi
FIGURE 1 The CONSORT flow diagram for cluster-randomised evaluation (WP4a)	30
FIGURE 2 Economic evaluation: risky prescribing model structure (WP4b)	36
FIGURE 3 Economic evaluation: risky prescribing submodel – CKD (WP4b)	37
FIGURE 4 Economic evaluation: BP control model structure (WP4b)	38
FIGURE 5 Economic evaluation: composite indicator of risky prescribing – scatterplot of simulated incremental cost and QALY (WP4b)	40
FIGURE 6 Economic evaluation: composite indicator of risky prescribing – CEAC	40
FIGURE 7 Economic evaluation: risky prescribing indicator 4 (prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of gastroprotection) – scatterplot of simulated incremental cost and QALY (WP4b)	41
FIGURE 8 Economic evaluation: risky prescribing indicator 4 (prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of gastroprotection) – CEAC (WP4b)	41
FIGURE 9 Economic evaluation: risky prescribing indicator 2 (prescribing of traditional oral NSAID in patients aged ≥ 75 years without coprescription of gastroprotection) – scatterplot of simulated incremental cost and QALY (WP4b)	42
FIGURE 10 Economic evaluation: risky prescribing indicator 2 (prescribing of traditional oral NSAID in patients aged ≥ 75 years without coprescription of gastroprotection) – CEAC (WP4b)	42
FIGURE 11 Economic evaluation: composite indicator of risky prescribing – CEAC for three sets of fixed intervention costs (WP4b)	43
FIGURE 12 Economic evaluation: composite indicator of BP control – scatterplot of simulated incremental cost and QALY (WP4b)	43
FIGURE 13 Economic evaluation: composite indicator of BP control – CEAC (WP4b)	44
FIGURE 14 Economic evaluation: BP control indicator 5 (achievement of BP $< 140/90$ mmHg in patients aged < 80 years with peripheral arterial disease) – scatterplot of simulated incremental cost and QALY (WP4b)	44
FIGURE 15 Economic evaluation: BP control indicator 5 (achievement of BP $< 140/90$ mmHg in patients aged < 80 years with peripheral arterial disease) – CEAC (WP4b)	45
FIGURE 16 Economic evaluation: cost-effectiveness plane for diabetes control implementation package (WP4b)	46

LIST OF FIGURES

FIGURE 17 Cluster-randomised evaluation: distribution of delivery of outreach visits across the intervention period	51
FIGURE 18 Process evaluation: intervention logic model (WP5)	57
FIGURE 19 Process evaluation: example of implementation process model for Treetop Practice (WP5)	58

List of boxes

BOX 1 Summary of 'high-impact' QIs agreed following consensus process (work package 1)	4
BOX 2 Illustrative denominator and numerator information for two 'high-impact' QIs (WP1)	77

List of supplementary material

Report Supplementary Material 1 Tables and figures

Supplementary material can be found on the NIHR Journals Library report page (<https://doi.org/10.3310/pgfar08040>).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

A&F	audit and feedback	NICE	National Institute for Health and Care Excellence
AF	atrial fibrillation		
ASPIRE	Action to Support Practices Implementing Research Evidence	NIHR	National Institute for Health Research
		NPT	normalisation process theory
BCT	behaviour change technique	NSAID	non-steroidal anti-inflammatory drug
BP	blood pressure		
CCG	Clinical Commissioning Group	OR	odds ratio
CEAC	cost-effectiveness acceptability curve	PPI	patient and public involvement
CHA ₂ DS ₂ -VASc	Congestive heart failure, hypertension, age > 75 years, diabetes mellitus, stroke and vascular disease	PSA	probability sensitivity analysis
		QALY	quality-adjusted life-year
		QI	quality indicator
CHADS ₂	Congestive heart failure, hypertension, age > 75 years, diabetes mellitus and prior stroke	QOF	Quality and Outcomes Framework
		SD	standard deviation
		SE	standard error
CI	confidence interval	TDF	theoretical domains framework
CKD	chronic kidney disease		
GP	general practitioner	TIA	transient ischaemic attack
HbA _{1c}	glycated haemoglobin	UKPDS	United Kingdom Prospective Diabetes Study
ICER	incremental cost-effectiveness ratio	WP	work package
IMD	Index of Multiple Deprivation	WTP	willingness to pay
MI	myocardial infarction		

Plain English summary

Clinical research identifies ways to improve patient care that can help people live longer and better. However, recommendations from research do not always become part of everyday patient care. Providing health-care staff and services with recommendations from clinical guidelines to inform clinical decision-making is often not enough by itself to change how care is delivered. Furthermore, general practice, where 9 out of 10 NHS patient contacts occur, struggles to apply many different guidelines within limited time and resources.

We developed and tested an improvement package to help general practices adapt their care according to one of four guideline recommendations.

We identified recommendations that were most likely to make a real difference to patient care. We measured how closely practices followed these recommendations and found that practices varied a lot. We adapted the package for use in one of four topics: diabetes control, blood pressure control, stroke prevention and 'risky prescribing' (to reduce particular combinations of medicines that may cause harm). We used the play of chance to assign general practices to one package for each topic. We found that only the risky prescribing package improved patient care and provided good value for money. We also found that the improvement packages were often not delivered or used as we had intended, and that professionals found it easier to change some types of clinical practice (e.g. prescribing) than others.

Our findings will inform practical guidance to help general practices improve patient care.

Scientific summary

Background

Clinical evidence that can improve patient outcomes does not reliably find its way into everyday care. The gap between evidence and practice limits the health, social and economic impacts of clinical research. Dissemination of evidence-based practice via clinical guidelines is necessary but seldom sufficient by itself to ensure implementation. The general practice context presents particular challenges – especially given limited practice organisational capacity, increasing workload and complexity of care, and competing priorities. Furthermore, many implementation studies focus on one condition. This limits confidence in applying research findings; it is uncertain how an intervention developed for one clinical condition will work for another. It is also impracticable and inefficient to invent and evaluate an implementation strategy for every new guideline.

Aim and objectives

We aimed to develop and evaluate an implementation package that could be adapted to support the uptake of a range of guideline recommendations and sustainably integrate within general practice systems and resources. We undertook this through five linked work packages, which are also summarised in *Figure a*.

Work package 1 identified and developed ‘high-impact’ quality indicators where a measurable change in clinical practice can lead to significant patient benefit.

Work package 2 measured and analysed levels of adherence to high-impact indicators.

Work package 3 developed an implementation package that is adaptable to target different indicators. This included theory-guided interviews to understand adherence to multiple indicators in primary care (work package 3a) and systematic, stakeholder-guided intervention development (work package 3b).

Work package 4 evaluated the effects and cost-effectiveness of the adapted implementation package in targeting the implementation of high-impact indicators. This included a cluster-randomised evaluation (work package 4a) and economic modelling (work package 4b).

Work package 5 comprised an in-depth process evaluation examining implementation package delivery and mechanisms of action.

Work package 1: identifying and developing ‘high-impact’ quality indicators

We reviewed existing national clinical guidelines and quality indicators and used a four-stage consensus development process to derive a set of ‘high-impact’ indicators relevant to primary care. Prioritisation criteria included burden of illness, potential for significant patient benefit, scope for improvement on current levels of achievement, the extent to which following a recommendation is directly within the control of individual practice teams, and the feasibility of measurement using routinely collected data. We screened 2365 recommendations to produce a shortlist of 102. These were considered by a consensus panel of 11 members, including primary care professionals, commissioners and patient representatives. We derived a set of 18 indicators (five single, 13 composites – comprising 2–9 individual recommendations) for field testing.

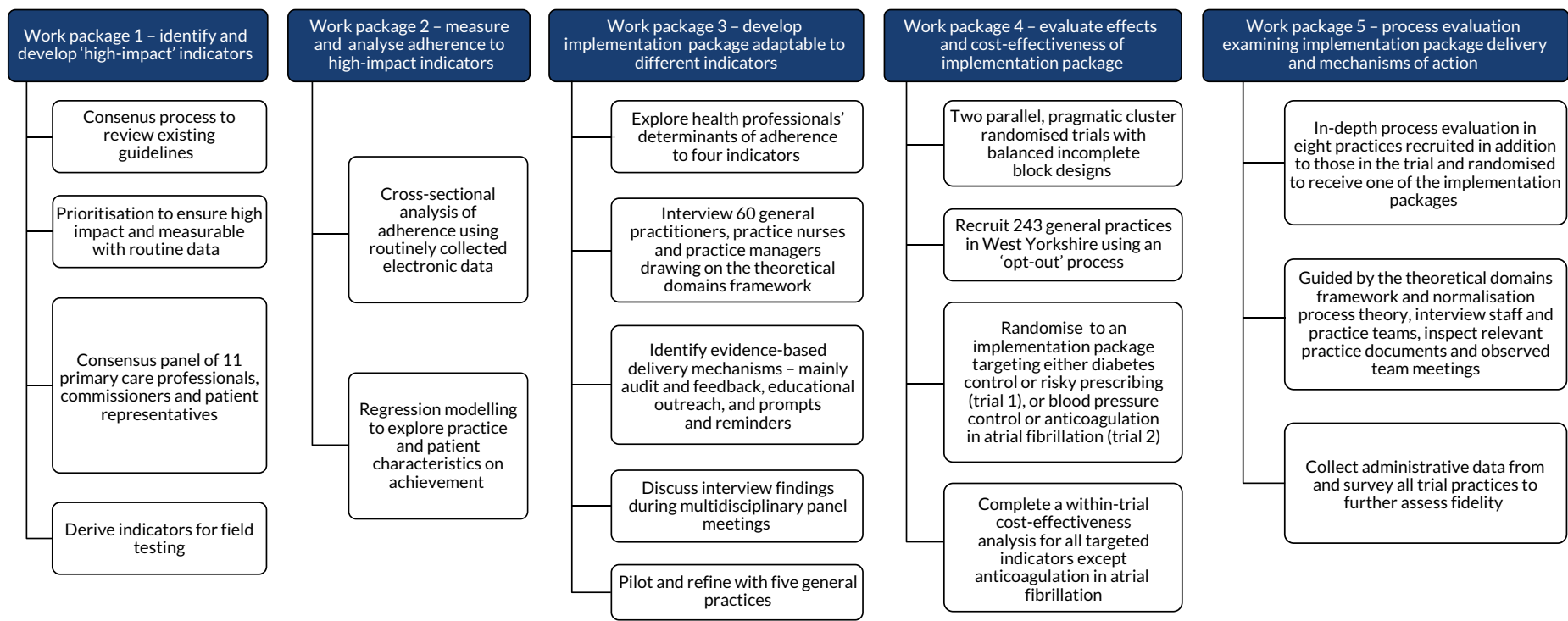


FIGURE a Research pathway diagram.

Work package 2: analysing adherence to 'high-impact' indicators

We undertook a cross-sectional analysis of adherence to seven indicators using routinely collected electronic data from a sample of 89 general practices in West Yorkshire. The indicators spanned processes and intermediate clinical outcomes of care related to diabetes, hypertension, atrial fibrillation, myocardial infarction, chronic kidney disease and 'risky prescribing' combinations. Regression modelling explored the impact of practice and patient characteristics on indicator achievement.

Median practice achievement of indicators ranged from 43.2% (diabetes control) to 72.2% (blood pressure control in chronic kidney disease). Considerable between-practice variation existed for all indicators: the absolute difference between the highest and lowest performing practices was 26.3% for risky prescribing and 100% for anticoagulation in atrial fibrillation. Odds ratios associated with the random effects for practices emphasised this; there was a greater than 10-fold difference in the likelihood of achieving the hypertension indicator between the lowest and highest performing practices (odds ratio range 0.50–5.24). Patient but not practice characteristics were modestly and consistently associated with indicator achievement, particularly age, gender and comorbidity.

Practice and patient characteristics partly accounted for marked inappropriate variations in practice. This may, in part, reflect the limitations of using routinely collected data but it is also likely that much of the remaining variation is attributable to differences in clinical and organisational behaviour.

Work package 3: developing an implementation package

We initially explored health professionals' perceived determinants of adherence to four indicators: achievement of recommended treatment targets for all of glycated haemoglobin, blood pressure and cholesterol in type 2 diabetes; avoidance of risky prescribing involving non-steroidal anti-inflammatory and anti-platelet drugs; achievement of anticoagulant prescribing for stroke prevention in atrial fibrillation; and achievement of recommended blood pressure levels in hypertension. We interviewed 60 general practitioners, practice nurses and practice managers in West Yorkshire, drawing on the theoretical domains framework. Data were analysed using framework analysis. We examined the degree to which determinants were indicator specific or potentially generalisable across indicators.

Professional role and identity and environmental context and resources featured prominently across all indicators, whereas the importance of other domains, for example beliefs about consequences, social influences and knowledge, varied across indicators. We identified five meta-themes that broadly underlined the need to align the design of interventions targeting general practices with higher-level supports and broader contextual considerations. These included the perceived nature of the job and norms of practice; internal and external sources of support; communication pathways and interaction; meeting the needs of patients; and perceptions of indicators. Our findings suggested that it was feasible to develop interventions to promote the uptake of different evidence-based indicators that share common features while also including content-specific adaptations.

We next used a staged process to develop an implementation package adaptable to the four indicators. We identified evidence-based delivery mechanisms: mainly audit and feedback, educational outreach, and prompts and reminders. Research team members independently mapped determinants of adherence (from the above interviews) to candidate behaviour change techniques, resolving discrepancies by discussion. We discussed key interview findings during a series of multidisciplinary panel meetings, each involving 5 to 10 primary care professionals, quality improvement specialists and service commissioners, and prioritised likely determinants and intervention content. We piloted and refined components of the implementation packages with five general practices. We downgraded our original plan for more extensive piloting of the whole implementation package because we prioritised starting the trials 3 months earlier to coincide with data collection for the 2015–16 Quality and Outcomes Framework year.

Work package 4: evaluating effectiveness and cost-effectiveness

We undertook two parallel, pragmatic cluster-randomised trials with balanced incomplete block designs. We recruited general practices in West Yorkshire using an 'opt-out' process. We randomised practices to an implementation package targeting either diabetes control or risky prescribing (trial 1), or blood pressure control or anticoagulation in atrial fibrillation (trial 2). In trials 1 and 2, practices randomised to the implementation package for one indicator acted as control practices for the other implementation package, and vice versa. The implementation package was tailored to each indicator and included behaviour change techniques embedded within audit and feedback, educational outreach and computerised support. Respective primary end points assessed after 11 months comprised achievement of all recommended levels of glycated haemoglobin, blood pressure and cholesterol; risky prescribing levels; achievement of recommended blood pressure levels in people with hypertension or at increased risk of cardiovascular events related to other conditions (e.g. chronic kidney disease); and anticoagulation prescribing. An intention-to-treat analysis using two-level binary logistic regression models with patients nested within registered practices adjusted for both patient- and practice-level covariates. We completed a within-trial cost-effectiveness analysis for all targeted indicators except anticoagulation in atrial fibrillation.

We approached 243 eligible general practices and recruited 178. We randomised 80 practices to trial 1 (40 per arm) and 64 to trial 2 (32 per arm), with 34 no-intervention controls. The implementation package reduced risky prescribing (odds ratio 0.82, 97.5% confidence interval 0.67 to 0.99); the estimated number needed to treat to prevent one case of risky prescribing was 95. The package had no effect on other primary end points; the odds ratio for diabetes control compared with controls was 1.03 (97.5% confidence interval 0.89 to 1.18), for blood pressure control was 1.05 (97.5% confidence interval 0.96 to 1.16) and for anticoagulation in atrial fibrillation was 0.90 (97.5% confidence interval 0.75 to 1.09).

The risky prescribing implementation package was more expensive and on average more effective than usual practice. This yielded an incremental cost-effectiveness ratio of £2337 per incremental quality-adjusted life-year, falling below the National Institute for Health and Care Excellence preferred threshold of £20,000–30,000 per quality-adjusted life-year. Over 75% of simulations conducted for the probabilistic sensitivity analysis fell below this threshold. The blood pressure implementation package was also more expensive and non-significantly more effective than usual practice, yielding an incremental cost-effectiveness ratio of £3954. However, the uncertainty around this result was considerable and the intervention incremental cost-effectiveness ratio had just over a 50% chance of falling below the £20,000 per quality-adjusted life-year threshold in the sensitivity analysis. Modelling indicated that the diabetes implementation package was unlikely to have favourable cost per quality-adjusted life-year estimates.

Work package 5: process evaluation

We conducted an in-depth process evaluation in eight practices recruited in addition to those in the trial and randomised to receive one of the implementation packages (two practices per package). Guided by the theoretical domains framework and normalisation process theory, we interviewed individual staff and practice teams, inspected relevant practice documents (e.g. meeting notes and policies) and observed team meetings. We also collected administrative data from and surveyed all trial practices to further assess fidelity. We analysed interview and observational data using a framework approach; this included constructs from the theoretical domains framework to compare planned versus actual intervention content and normalisation process theory to understand individual and group implementation processes. We compared trial fidelity data with in-depth practice case narratives to help to explain trial findings.

We had set out to devise an implementation package that could be adapted to target a range of high-impact indicators and that could also be delivered within existing resources and ways of working in general practice. Our process evaluation offers three main explanations for its limited impact. First, we observed losses in fidelity from delivery through to enrolment and enactment by practice teams; for example, although all practices received feedback reports, under half (67; 46.5%) accepted outreach visits. Second, the type and scale of targeted behaviour changes varied by indicator, such that practice teams may have perceived and exerted greater control over goal-setting and action-planning for risky prescribing. In some cases, receiving quarterly feedback reports that drew attention to perceived insufficient progress towards achieving targets (particularly in the diabetes and blood pressure control practices) seemed to activate a negative feedback loop and demotivate staff. Third, we had sought to ensure that targeted indicators were well aligned with existing clinical priorities; this inadvertently resulted in the implementation package being insufficiently differentiated from other improvement initiatives.

Patient and public involvement

Our patient and public involvement panel comprised nine people from diverse ethnic, occupational and social backgrounds and with considerable collective lay experience in commissioning and governance of health care, national clinical audits, patient advocacy and community development. The panel met quarterly and contributed to all work packages and wider stakeholder events. In addition, we undertook further work on the role of patient and public involvement in implementation research.

Patient and public involvement contributions to Action to Support Practices Implementing Research Evidence

The panel played an integral role in the development and conduct of the Action to Support Practices Implementing Research Evidence (ASPIRE) programme. Specific contributions included patient representation in consensus processes (e.g. the selection of research areas; work package 1), endorsement of an opt-out recruitment process (work package 2 and work package 4), feedback on the interpretation of emerging data (all) and assistance with relationship-building and dissemination.

The role of patient and public involvement in implementation research

Patient and public involvement is generally an essential requirement for research funding. As an implementation research programme, ASPIRE generally focused on health professionals' behaviour. This is in contrast to clinical research, which generally focuses on patients. Discussions with our panel identified uncertainty about the role of patient and public involvement in implementation research and we decided to explore this issue further to inform relevant good practice guidance. Via a structured consensus process, our panel considered and rated 21 potential patient and public involvement roles in research. There were more disagreements relating to patient and public involvement roles in implementation research than in clinical research. The work informed a framework to guide the planning, conduct and reporting of patient and public involvement in implementation research.

Conclusions

Our highly pragmatic and rigorous evaluation indicates the value of an implementation package targeting risky prescribing, given predictable population reductions in avoidable morbidity, deaths and hospital admissions. However, in broad terms, an adapted 'one-size-fits-all' approach did not work, with no improvement for other targeted indicators.

Our programme had several limitations. It was set in one geographical area; however, practice and patient population characteristics are otherwise likely to be sufficiently diverse and typical to enhance generalisability. We used an 'opt-out' approach to recruit general practices to the randomised trials. Subsequently, our trial practices may not have engaged with the implementation package as much as if

they had actively volunteered. However, this approach may also have ensured the applicability of our findings to 'real-world' quality improvement initiatives targeting all practices in a given locality.

There are challenges in designing implementation strategies that are sufficiently robust to bring about change in the face of difficult clinical contexts and likely losses to fidelity. Despite our systematic intervention development, we now believe that we could have conducted more feasibility and 'stress-testing' work prior to rolling out interventions within a definitive evaluation – and recommend this to others. Our programme has led onto other work, adapting our audit and feedback approach for other priorities and evaluating different ways of delivering feedback to improve patient care. We are also producing practical, evidence-based guidance and supporting materials to promote the implementation of National Institute for Health and Care Excellence guidance in general practice.

Implications for practice

Selecting priorities for implementation may be facilitated by a realistic appraisal of the likelihood of being able to bring about change for the targeted clinical practice, ideally ensuring that any goals for change are within the control of the professionals and patients who need to change behaviour. Specifically, we demonstrated a reduction in risky prescribing by a strategy exploiting routinely available data and involving repeated audit and feedback accompanied by persuasive messages, realistic goal-setting and action-planning. Implementation strategies may have small to modest effects but such effects can translate into worthwhile population health gains.

Recommendations for research

We recommend maximising feasibility and 'stress-testing' work prior to rolling out interventions within a definitive, pragmatic trial. We specifically recommend further implementation research addressing type 2 diabetes control, blood pressure control and anticoagulation for atrial fibrillation that builds on our experience and the wider body of research literature. Concerted strategies that target system, organisational and patient levels as well as general practices may be required to bring about significant change.

Trial registration

This trial is registered as Current Controlled Trials ISRCTN91989345.

Funding

This project was funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research programme and will be published in full in *Programme Grants for Applied Research*; Vol. 8, No. 4. See the NIHR Journals Library website for further project information.

SYNOPSIS

Work package 1: developing 'high-impact' quality indicators for primary care

Parts of this section are reproduced from Rushforth *et al.*¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Clinical evidence that can cut avoidable deaths and enhance quality of life does not reliably find its way into everyday patient care. The translation of evidence into practice is unpredictable and can be a slow and haphazard process.² This gap between evidence and practice is a strategically important problem for policy-makers, health-care systems and research funders because it limits the health, social and economic impacts of clinical research.³

The primary care context presents particular implementation challenges – given growing demand, increasing complexity of care and limited workforce capacity, and against a background of continual organisational reconfigurations and the dispersed and independent nature of practices.^{4–7} An international review of quality-of-care studies from primary care concluded 'In almost all studies reviewed the quality of care did not attain acceptable standards of practice'.⁸ A number of initiatives in the UK have aimed to increase implementation of effective practice in primary care. These include the development and dissemination of evidence-based clinical guidelines by the National Institute for Health and Care Excellence (NICE)⁹ and financial incentives to reward adherence to performance indicators set out in the Quality and Outcomes Framework (QOF).¹⁰

Measuring adherence to recommended practice is a cornerstone of any strategy to improve quality of care. Measurement is required to identify inappropriate variations in practice, target improvement efforts and monitor their impact. The development of quality indicators (QIs) from clinical guidelines offers a way to assess adherence to recommended practice.^{11–13} Formal consensus methods are generally used both to prioritise clinical guideline recommendations suitable for indicator development and to develop valid and reliable indicators.^{13–15}

Several challenges and considerations need to be balanced in developing indicators:

- Indicators developed solely by expert panels may be 'unoperationalisable, unreliable, too rare to be useful, or too hard to extract reliably'.¹⁶
- Methods requiring manual data extraction are resource intensive.
- The utility of routinely collected data drawn from existing schemes, such as QOF, is limited by incomplete coverage of health care.¹⁷
- Health-care process indicators need a strong evidence base showing that the care process leads to improved outcomes.¹⁸
- Indicators focusing on processes of care rather than health outcomes may not help overcome therapeutic inertia (i.e. the failure to intensify treatment in patients with an abnormal clinical measurement).¹⁹
- Health outcome indicators are subject to higher 'noise-to-signal' ratios, whereby a range of factors beyond professional practice influence outcomes.^{20–22}

Improvement strategies in primary care also need to take account of efficiency. Implementation studies generally focus on one clinical condition. This has advantages, for example so that an intervention to promote better detection of hypertension can complement another to improve the treatment of detected hypertension. However, the impact and generalisability of such studies is limited in a number of ways:

- Only a minority of single-issue guideline recommendations are relevant to primary care and sufficiently clinically important to justify concerted implementation and provide an acceptable return on investment.
- Many important clinical practice recommendations are not directly amenable to measurement.
- There are risks of encountering 'ceiling effects' when adherence to a given recommendation has reached a point beyond which it is difficult to improve practice further.

General practitioners (GPs) need to contend with a large number of implicitly competing indicators. We used a structured multistage consensus process and field testing to prioritise clinical practice recommendations and develop a set of 'high-impact' indicators that could be measured using routinely recorded data and, if implemented, yield significant patient and population benefits.

Methods and results

Stage 1: initial screening of candidate recommendations

We identified candidate recommendations and indicators from 147 NICE clinical guidelines (published December 2002–June 2012), 19 NICE quality standards (June 2010–June 2012) and 95 QOF clinical indicators (extracted June 2012). One researcher (BR) screened titles and summaries of NICE guidelines and quality standards for relevance to primary care; 20 were excluded as they were exclusively related to secondary care. We excluded a further 20 that had been superseded by a more recent update. We excluded four QOF indicators mainly related to secondary care. Together, these sources yielded a total of 2365 candidate recommendations.

Two clinical researchers (BR and RF) then independently screened candidate recommendations, discarding those judged irrelevant to primary care or not measurable using routine data. We grouped clearly linked sets of recommendations to form 'composite' recommendations (e.g. the nine recommended processes of care for patients with type 2 diabetes).²³ We resolved disagreements through discussion. The final 'longlist' of 102 candidate recommendations comprised 56 single and 46 composite recommendations (additional files published with Rushforth *et al.*¹).

Stage 2: online shortlisting by consensus panel

We used a modified RAND consensus process.¹⁴ We convened an 11-member multidisciplinary consensus panel, which comprised five GPs (including two with responsibilities for commissioning services), a practice nurse, a practice manager, a consultant clinical advisor from NICE, a health informatics specialist and two patient representatives. We deliberately weighted the panel towards professionals who would typically act on clinical practice recommendations, recognising that a number of judgements required an in-depth, tacit understanding of the day-to-day realities of clinical practice. The panel was limited to 11 members as there are only marginal gains in reliability beyond this number.¹⁴

Each panellist independently rated all 102 recommendations via an online survey according to three criteria: burden of illness (e.g. prevalence, severity, costs), potential for significant patient benefit (e.g. longevity, quality of life, safety of care) and scope for improvement on current levels of adherence (e.g. from perceived current low levels or large variations). All ratings were completed on a 9-point Likert scale (where 1 is low and 9 is high according to their perceptions of current practice) with 'don't know' options available.

Panel ratings of recommendations were generally high for patient burden [mean 'median' score of 7.6; standard deviation (SD) 0.68] and potential for patient benefit (7.8; SD 0.81), with lower scores for scope for improvement (5.00; SD 0.88). We excluded 18 recommendations from further review at this stage because they scored ≤ 4 on scope for improvement (indicating that the panel perceived adherence to these recommendations to be relatively good).

A second online survey contained the top 62 recommendations (31 single and 31 composites) based on the highest aggregate rankings. Panellists rated shortlisted recommendations using three further criteria: the feasibility of measuring adherence (e.g. from routinely collected clinical data); the extent to which following a recommendation is directly within the control of practice teams or individual professionals; and the likelihood of cost savings without patient harm (all assessed on 1–9 Likert scales). Thresholds for disagreement were defined in advance as at least three panellists scoring a recommendation 1–3 on a particular criterion, and at least three scoring it 7–9. The panel disagreed on a total of 22 (11.8%) ratings; 20 disagreements concerned the feasibility of measuring adherence and two concerned the extent to which following a recommendation was directly within the control of practice teams or professionals.

Stage 3: face-to-face consensus panel meeting

The panel met for a facilitated, structured discussion, led by an experienced researcher. We presented summaries of the evidence for each recommendation, clarified any aspects of recommendations and discussed reasons for low or high rankings with a view to reaching, but not forcing, consensus. Panellists independently re-rated each recommendation immediately after discussing each recommendation. Following the panel meeting, there were disagreements for 12 (6.4%) ratings. Across the 62 shortlisted recommendations, the mean 'median' ratings were 6.8 (SD 1.57) for the feasibility of measuring adherence, 7.2 (SD 0.76) for the extent to which following a recommendation is directly within the control of individual practice teams or professionals and 7.3 (SD 0.73) for the likelihood of cost savings without patient harm. This process produced a ranked list of 50 recommendations for which consensus was achieved.

Stage 4: informal sense-checking

We added a sense-checking exercise to guide the final selection of recommendations, aiming for a list of around 20 that could be taken forward to the next stage of the programme. During the consensus and ranking exercise, we were struck by some unexpected anomalous rankings that appeared to lack face validity when considered against our rating criteria. For example, we doubted the feasibility of using routinely available data to measure adherence on recommended secondary prevention following myocardial infarction (MI), which had made the top 20 list after the panel rating. Current and past activity levels and preferences should be considered when advising on physical activity. An appropriately qualified health professional can help tailor advice on the benefits of exercise.²⁴ We also wanted to ensure that identified recommendations would be consistent with local priorities although their measurement was unlikely to face ceiling effects given known national and local initiatives.

We therefore identified a convenience sample of four GP commissioning leads and six academic GPs that we had existing working relationships with and who had practical experience of measuring primary care outcomes. We e-mailed and asked them to review the full ranked list of recommendations from the consensus panel. We invited them to select between 5 and 10 recommendations that they considered would best meet our aims and highlight any that they considered problematic to target. We then collated their selections and written comments. During this process, we amalgamated two similar recommendations (concerning initiation of insulin in type 2 diabetes) and replaced one recommendation concerning prescribing non-steroid anti-inflammatory drugs (NSAIDs) with a composite recommendation on risky prescribing.²⁵ Comments from this sense-checking exercise centred on concerns regarding perceived likelihoods of ceiling effects, difficulties in measurement or recommendations being outside the immediate control of the primary care team. This process produced 18 recommendations (Box 1), 11 of which had been ranked in the top 20 by our panel. These mainly covered chronic disease management and cardiovascular disease.

Stage 5: field testing of indicators

We assessed the extent to which it was feasible to operationalise the QIs in a sample of general practices [sampling described under work package (WP) 2]. Search algorithms were generated by a clinical researcher (BR) and a primary care data analyst. We applied and iteratively refined the algorithms with input from two external GP advisors. Appendix 1, Box 2, illustrates two examples. The full set of SystmOne (The Phoenix Partnership, Leeds, UK) searches for each of the 18 recommendations is available in Rushforth *et al.*¹

BOX 1 Summary of 'high-impact' QIs agreed following consensus process (work package 1)

1. Smoking: the percentage of patients in high-risk groups whose notes record smoking status and the offer of support and treatment within the preceding 15 months (composite).
2. COPD: diagnosis of COPD through use of spirometry and chest radiograph (composite).
3. CKD: the percentage of patients on the CKD register with hypertension and proteinuria who are treated with an angiotensin-converting-enzyme inhibitor or angiotensin receptor blocker.
4. CKD: measurement of blood pressure, urinary protein excretion and lifestyle advice (composite).
5. CKD: blood pressure and urinary protein excretion targets, and appropriate drug therapy (composite).
6. MI: all patients who have had an acute MI should be offered specific combination drug treatment.
7. Chronic heart failure: measurement of serum natriuretic peptides and referral where appropriate (composite).
8. AF: recommendations concerning use of anticoagulants in AF (composite).
9. Hypertension: blood pressure targets in those aged under/over 80 years (composite).
10. Hypertension: lifestyle advice and monitoring of cholesterol and urinary protein excretion (composite).
11. Type 2 diabetes: nine annual processes of care (i.e. measurement of blood pressure, lipids, renal function, urine ACR, glycaemic control, BMI, smoking status, plus foot and eye checks) (composite).
12. Type 2 diabetes: integrate dietary advice with a personalised diabetes management plan.
13. Type 2 diabetes: cardiovascular risk assessment and subsequent statin therapy when indicated.
14. Type 2 diabetes: achievement of target levels for blood pressure, cholesterol and glycaemic control (composite).
15. Type 2 diabetes: for a person on dual therapy who is markedly hyperglycaemic, consider starting insulin therapy in preference to adding other drugs to control blood glucose.
16. Diabetes mellitus: the percentage of patients with diabetes in whom the last blood pressure is $\leq 140/80$ mmHg.
17. Risky prescribing: indicators focusing on avoiding adverse gastrointestinal, renal and cardiac effects of NSAIDs and anti-platelet drugs (composite).
18. Depression in adults: recommendations concerning severity-appropriate treatment of depression (composite).

ACR, albumin–creatinine ratio; AF, atrial fibrillation; BMI, body mass index; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease. Reproduced from Rushforth *et al.*¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Discussion

We developed high-impact indicators that can be measured using routinely collected data in primary care. Our development process required considerable filtering of existing guidelines and depends on the availability of routinely recorded data. Our 18 indicators were drawn from 2365 recommendations and indicators; earlier research attempting similar work also found this to be a labour-intensive process with a limited 'yield'.²⁶ Our indicators overlapped substantially with existing primary care QI sets^{17,27,28} that largely focus on long-term conditions, such as diabetes and cardiovascular disease.

We highlight five limitations. First, our process lost a degree of transparency through the addition of a less formal 'sense-checking' stage. The need to add this stage somewhat highlights a relative failure of our preceding consensus process to scrutinise the candidate indicators. Panels developing indicators may tend to overestimate the feasibility of data collection.²⁹ Second, our indicator set is skewed towards biomarkers (e.g. glycaemic control in diabetes) that are used for chronic disease monitoring. We recognise the risk of marginalising holistic medical care through focusing attention on what is measurable and what is not necessarily important to patients or physicians.³⁰ However, as well as including patient representatives in our consensus process, we also sought to maintain a focus on recommendations supported by evidence of benefits for patient and population outcomes (e.g. smoking cessation). Third, our approach to indicator development prioritised those associated with higher population burdens of illness; this discounts rare diseases for which appropriate care could make a major difference to individual outcomes.³¹ We recognise that we made a trade-off. Fourth, we did not directly assess the reliability of data recording. However, our measures were mostly derived from data that either had been through reliability checks during piloting or were QOF indicators.³² Fifth, the detailed operationalisation of our indicators is only relevant to UK primary care. Nevertheless, their evidence base and basic structures may be transferable to similar primary care settings.

Conclusion

We developed 18 high-impact QIs that can be measured using routinely collected data. Our methods were more iterative and required more judgement than originally planned, especially considering our additional sense-checking stage and refinements following field testing.

Work package 2: variations in achievement of evidence-based, high-impact quality indicators in general practice

Parts of this section are reproduced from Willis *et al.*³³ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Background

There is well-documented variation in the delivery of primary care in the UK and internationally.^{8,34–39} 'Big data' offer considerable promise in assessing population health-care and researching variations.^{40,41} For example, the *NHS Atlas of Variation in Healthcare*⁴² illustrates large geographical variations in care across England for several clinical conditions, including diabetes, stroke and cancer. The magnitude of these differences cannot be easily explained by population and case mix factors; much variation is likely to be attributable to 'idiosyncratic practices of clinicians and of healthcare organisations.' (Reproduced from NHS Right Care.⁴² Contains public sector information licensed under the Open Government Licence v3.0). Such variations can, therefore, be considered to be inappropriate if patients are not receiving recommended care.

Gaps between recommended and actual care can have substantial implications; for example, in England an estimated 7000 strokes per year could be prevented and 2100 lives saved from increased adherence to guidelines on the management of atrial fibrillation (AF).⁴³ Clearly, quality improvement initiatives cannot focus on all clinical practice recommendations at once; there is a need to identify and prioritise those with the potential for the most patient and population benefit. We examined levels of adherence to selected high-impact QIs, developed in WP1, and assessed the extent to which variations in achievement could be explained by practice and patient characteristics using routinely collected data.

Methods

Study design and setting

We conducted a cross-sectional analysis of achievement against QIs using routinely collected, electronic data from randomly sampled general practices in West Yorkshire.

Participants

Of 334 general practices in West Yorkshire at the time of the study, 272 used the SystmOne clinical information system which permitted centralised data extraction. We sampled randomly from this group and stratified them by the then configured five NHS primary care trusts. We assumed a 30% decline rate based on earlier work⁴⁴ and initially asked 78 practices to 'opt in' to the sharing of anonymised data to achieve a sample of 60 (see *Sample size*). After receiving several declines and a small number of acceptances, we approached and sampled an additional 36 practices, making 114 in total. At the same time, the local NHS Research Ethics Committee granted permission to change to 'opt-out' recruitment to reduce selection bias by facilitating general practices' agreement to share anonymised patient data.

Variables

We selected seven of the QIs developed in WP1 based on likely scope for improvement and amenability to change (see table 1 of Willis *et al.*³³). Four indicators focused on processes of care (e.g. prescribing) and three focused on clinical outcomes (achievement of treatment goals).

We examined patient-level demographic variables (age, gender and ethnicity) and comorbidities as recorded for QOF disease registers. Practice-level variables comprised the number of GP partners (a proxy for practice size), the number of salaried GPs and training status. We used practice-level Index of Multiple Deprivation (IMD) scores. We used overall QOF clinical domain (2012–13) achievement as a proxy measure for overall quality of care. Patient data were remotely extracted and anonymised before transfer. We obtained data from *National General Practice Profiles*⁴⁵ for two further practice-level variables: patient satisfaction (the proportion recommending the practice to others) and accessibility (the proportion reporting being able to speak with a GP or nurse within 48 hours of approach).

Data analysis

For each QI, we assessed the proportion of cases with documented receipt of appropriate care or attainment of treatment goals. Denominators were eligible patients, identified by diagnostic codes and prescribing records. Numerators were patients with evidence of a clinical intervention offered or received, or meeting defined treatment goals. In assessing the impact of practice and patient characteristics, we initially calculated unadjusted odds ratios (ORs) before adjusting for other variables associated with outcomes.

Data for almost all patients were complete. Data on age were missing from a small proportion (< 1%) of patients and excluded from analysis.

Sample size

Effect size calculations informed a recruitment target of 60 practices. With seven covariates, and a large effect size (defined by a difference of at least 0.8 SDs),⁴⁶ 60 practices would provide 94% power.

Results

Participants

Eighty-nine practices (78.1% of those approached) shared patient data. Practices declining participation only differed from participating practices in having a smaller mean number of GPs (5 vs. 3.6; $p = 0.05$). The total number of patients for each indicator denominator ranged from 4773 (anticoagulation in AF) to 77,587 [blood pressure (BP) control]. Willis *et al.*³³ summarises patient demography (see table 2 of Willis *et al.*³³). Practice size was indicated by the number of practice partners (mean 3.7, SD 2.3) and

salaried GPs (mean 1.3, SD 1.8). Mean practice-aggregated IMD score was 31.2 (SD 11.9, approximating to the highest quarter of deprivation). Mean QOF 2012–13 performance for clinical domains across practices was 637.4 (SD 27.6; approximating to the national mean⁴⁷) and 20.2% were training practices.

Achievement of indicators

Median practice achievement of the indicators ranged from 43.2% (range 20.8–66.2%) for diabetes outcomes to 74.2% (range 50.7–100%) for BP control in chronic kidney disease (CKD) (see tables 3 and 4 in Willis *et al.*).³³ Median achievement of the risky prescribing indicator was 8.7%, although lower scores were indicative of fewer instances of risky prescribing and were, therefore, desirable. Considerable between-practice variation in achievement existed for all indicators: the difference between the highest and the lowest achievers was 26.3% for risky prescribing and 100% for anticoagulation in AF and BP control in CKD.³³

Associations with achievement

The range of ORs associated with the random effects for practices demonstrate that the likelihood of achieving a specific indicator varied substantially as a consequence of the practice at which a patient was registered (see tables 3 and 4 in Willis *et al.*).³³ These ORs were typically of a much greater magnitude than those for other variables, demonstrating strong practice effects. For process indicators, the impact of the practice attended was most pronounced for risky prescribing, with a sevenfold difference between the lowest and highest performing practices (OR range 0.40–3.51). Practice effects were least apparent for secondary prevention of MI (OR range 0.70–1.42). There were also sizeable practice effects for outcome indicators. For the achievement of target BP values in hypertension there was a greater than 10-fold difference between the highest and lowest performing practices (OR range 0.50–5.24) and a fourfold difference for diabetes control (OR range 0.51–2.05). Practice effects were less marked for the achievement of BP targets in CKD (OR range 0.54–1.60). Across the seven indicators, statistically significant associations were identified more frequently with patient than with practice characteristics (see tables 3 and 4 in Willis *et al.*).³³ The amount of variance explained by these variables, however, was relatively low; practice characteristics explained less than 8% of variance across all seven models. Variance due to patient ethnicity typically explained a small amount of variance in achievement (< 10% of the variation due to practices).

Process indicators

Diabetes processes of care

Males were more likely to receive all nine of the recommended processes of care in diabetes than females [adjusted OR 1.24, 95% confidence interval (CI) 1.17 to 1.30]. Relative to younger patients, receipt was more likely in each of the age groups > 40 years old: 40–59 years (adjusted OR 1.52, 95% CI 1.33 to 1.73), 60–79 years (adjusted OR 2.07, 95% CI 1.81 to 2.36) and ≥ 80 years (adjusted OR 1.51, 95% CI 1.30 to 1.76). Indicator achievement was more likely in those with a greater number of comorbidities: compared with patients appearing on 0–3 QOF registers, the odds were higher for those on 4–5 QOF registers (adjusted OR 1.24, 95% CI 1.17 to 1.32) and on 6–13 QOF registers (adjusted OR 1.33, 95% CI 1.23 to 1.45).

Risky prescribing

Males were more likely to be prescribed at least one risky prescribing combination than females (adjusted OR 1.11, 95% CI 1.02 to 1.19). Risky prescribing was more likely in patients aged 40–59 years (adjusted OR 1.71, 95% CI 1.12 to 2.60) and 60–79 years (adjusted OR 1.95, 95% CI 1.26 to 2.96), but not in those aged ≥ 80 years, relative to patients < 40 years. Compared with patients with 0–3 comorbidities, risky prescribing was less likely in those on 4–5 (adjusted OR 0.81, 95% CI 0.74 to 0.88) and 6–11 (adjusted OR 0.56, 95% CI 0.51 to 0.62) QOF registers. Registration at a practice with a greater proportion of salaried GPs was associated with lower likelihood of risky prescribing (adjusted OR 0.76, 95% CI 0.61 to 0.94).

Anticoagulation in atrial fibrillation

Males were more likely than females to be prescribed anticoagulants (adjusted OR 1.27, 95% CI 1.12 to 1.44). Patients aged ≥ 80 years were less likely to be treated than those aged < 60 years (adjusted OR 0.62, 95% CI 0.43 to 0.89).

Secondary prevention of myocardial infarction

Males were more likely than females to be prescribed the four recommended medications (adjusted OR 1.12, 95% CI 1.02 to 1.23). Patients aged ≥ 80 years were less likely to be treated (adjusted OR 0.38, 95% CI 0.22 to 0.65) than those in the youngest quartile. Patients with higher levels of comorbidity, featuring on 6 or more QOF registers, were more likely to be treated than those on 0–3 registers (adjusted OR 0.83, 95% CI 0.73 to 0.94).

Outcome indicators***Diabetes control***

Achievement of all three target values for (glycated haemoglobin) HbA_{1c}, cholesterol and BP in diabetes was slightly higher in males than in females (adjusted OR 1.09, 95% CI 1.03 to 1.14). This likelihood increased with age, with all three age groups significantly more likely than patients aged < 40 years to achieve treatment goals (40–59 years: adjusted OR 1.28, 95% CI 1.10 to 1.47; 60–79 years: adjusted OR 2.55, 95% CI 2.21 to 2.94; ≥ 80 years: adjusted OR 2.91, 95% CI 2.48 to 3.40). Comorbidity levels were associated with indicator achievement: patients on 4–5 (adjusted OR 1.10, 95% CI 1.04 to 1.17) and 6–13 QOF registers (adjusted OR 1.31, 95% CI 1.21 to 1.42) were more likely to achieve treatment goals than those on 0–3 registers. Practices with better than average QOF performance (adjusted OR 1.19, 95% CI 1.02 to 1.39) and better reported accessibility (adjusted OR 1.18, 95% CI 1.02 to 1.38) were more likely to achieve this indicator.

Blood pressure control in hypertension

Males were less likely to achieve target BP values than females (adjusted OR 0.86, 95% CI 0.84 to 0.89). Achievement likelihood increased with age, with patients aged 60–79 years (adjusted OR 1.19, 95% CI 1.07 to 1.31) and particularly those aged ≥ 80 years (adjusted OR 3.34, 95% CI 2.99 to 3.74) more likely to achieve treatment goals than patients aged < 40 years. Patients with greater levels of comorbidity were more likely to achieve control than patients on 0–3 registers (3–4 QOF registers: adjusted OR 1.54, 95% CI 1.48 to 1.60; 5–13 registers: adjusted OR 2.32, 95% CI 2.20 to 2.44). Achievement was more likely in practices with better QOF performance (adjusted OR 1.24, 95% CI 1.06 to 1.46).

Blood pressure control in chronic kidney disease

Target achievement was less likely as the level of comorbidity increased, through 4–5 QOF registers (adjusted OR 0.88, 95% CI 0.81 to 0.95) and 6–13 QOF registers (adjusted OR 0.86, 95% CI 0.79 to 0.95).

Discussion

We found marked variations between general practices in the achievement of clinically important indicators. The odds of patients receiving recommended care or achieving recommended treatment targets varied between two- and over 10-fold by indicator according to the practice attended. These variations were partly explained by a range of routinely available practice and patient variables; it is likely that much remaining variation is related to clinical and organisational behaviours as well as unmeasured characteristics.

We highlight four study limitations. First, we considered quality of care from a single, technical perspective (i.e. achievement against selected clinical indicators). Nevertheless, the indicators were derived from a rigorous consensus process and we are confident of their importance to both clinicians and patients. Second, the study was limited to one geographical area and practices using one computerised patient record system. West Yorkshire has practice characteristics broadly similar to English averages and regional SystemOne coverage was high ($> 80\%$ of practices during the study). Furthermore, opt-out recruitment efficiently

enabled practice participation while avoiding biases associated with opt-in recruitment,^{44,48} strengthening generalisability. Third, combined indicators can mask varying performance between individual component indicators. Giving equal weighting to indicators can be contentious;⁴⁹ however, there is no single agreed method of combining indicators and our methods were similar to those used elsewhere (e.g. Steel *et al.*³⁸ and Levine *et al.*⁵⁰). Fourth, our rather crude measure of deprivation (practice-averaged IMD) did not consider deprivation at the individual patient level and may have masked differences in variation within practice populations. We had been unable to use patient-level data because of the risk of breaching confidentiality.

We highlight three implications for practice. First, the consistent, substantial variations between the practices that we observed are at similar levels to those identified almost two decades ago.⁸ Patients and, perhaps, clinicians and policy-makers might be surprised, if not concerned, to learn of such variations in the receipt of recommended care and achievement of treatment goals. Our findings suggest the continuing salience of inappropriate variations to policy and research agendas. Second, the associations between patient and practice variables and indicator achievement suggests the importance of clinical and organisational behaviours. There is some evidence to suggest that health-care professionals may believe that practice performance is predominantly influenced by local case mix and demography.⁵¹ Our findings suggest an interpretation that highlights the role of clinical and organisational behaviours. Third, the modest but significant associations between achievement and specific patient characteristics have implications for improvement strategies. Better performance on the diabetes, hypertension and risky prescribing indicators was associated with comorbidity. This suggests scope for focusing greater attention on patients who are (relatively) healthier and perhaps less likely to attend practices. Our analysis also identified associations between achievement and a range of further patient characteristics, which can guide targeting of improvement strategies.

Work package 3a: using the theoretical domains framework to understand adherence to multiple quality indicators in primary care

Parts of this section are reproduced from Lawton *et al.*⁵² This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Having demonstrated major variations in achievement of a set of high-impact QIs, we next explored primary care professionals' perceived determinants of adherence to these indicators. A wide variety of theories from behavioural science, economics and social marketing are available to understand clinical behaviour.⁵³ The theoretical domains framework (TDF) was specifically developed to identify determinants of professional behaviour change and includes knowledge, skills, beliefs about consequences, beliefs about capabilities, social influences, emotion, motivation/goals, professional role/identity, memory and decision processes, environmental context and resources, and action-planning.⁵⁴

We examined which TDF determinants were specific to indicators, thereby suggesting a need for indicator-specific tailoring of implementation strategies, and which were shared across all indicators, thereby suggesting the potential for incorporating common elements into implementation strategies across different indicators. In considering shared determinants that may represent wider contextual influences, we also looked for meta-themes that emerged when synthesising data from multiple indicators. This study is fully reported in Lawton *et al.*⁵²

Methods

Design and setting

We conducted semistructured interviews with primary care professionals in West Yorkshire, UK.

Indicator selection

We selected eight indicators from WP1 and WP2 for this interview study (one of which, advising on smoking cessation, we did not report under WP2 given doubts about its validity). Here, we focus on the four indicators subsequently targeted by our implementation package (Table 1):

1. risky prescribing, especially involving NSAIDs⁵⁵
2. treatment targets in type 2 diabetes⁵⁶
3. BP targets in treated hypertension⁵⁷
4. anticoagulation in AF.⁵⁸

TABLE 1 Quality indicators explored in qualitative interviews with primary care staff (WP3a)

Indicator topic	Indicator details
Risky prescribing	<p>Avoidance of the following prescribing combinations:</p> <ul style="list-style-type: none"> • prescribing of a traditional oral NSAID or low-dose aspirin in patients with a history of peptic ulceration without coprescription of a gastroprotective drug • prescribing of a traditional oral NSAID in patients aged ≥ 75 years without coprescription of a gastroprotective drug • prescribing of a traditional oral NSAID and aspirin in patients aged ≥ 65 years without coprescription of a gastroprotective drug • prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of a gastroprotective drug • prescribing of warfarin and a traditional oral NSAID without coprescription of a gastroprotective drug • prescribing of warfarin and low-dose aspirin or clopidogrel, without coprescription of a gastroprotective drug • prescribing an oral NSAID in patients with heart failure • prescribing an oral NSAID in patients prescribed both a diuretic and an angiotensin-converting-enzyme inhibitor/angiotensin receptor blocker • prescribing an oral NSAID in patients with CKD (stages 3, 4 and 5)
Treatment targets in type 2 diabetes	<p>Achievement of all three recommended levels:</p> <ul style="list-style-type: none"> • BP of $< 140/80$ mmHg (or $130/80$ mmHg if there is kidney, eye or cerebrovascular damage) • HbA_{1c} value of ≤ 59 mmol/mol • cholesterol level of ≤ 4.0 mmol/l in patients who are ≥ 40 years
BP targets in treated hypertension	<p>Aim for a target clinic BP of $< 140/90$ mmHg in people aged < 80 years with treated hypertension</p> <p>Aim for a target clinic BP of $< 150/90$ mmHg in people aged ≥ 80 years with treated hypertension</p>
Anticoagulation in AF	<p>In patients with AF who either are post stroke or have had a transient ischaemic attack:</p> <ul style="list-style-type: none"> • Warfarin should be administered as the most effective thromboprophylactic agent • Aspirin or dipyridamole should not be administered as thromboprophylactic agents unless indicated for the treatment of comorbidities or vascular disease <p>Those patients with AF in whom there is a record of a CHADS2 score of 1 should be offered antioagulation drug therapy or antiplatelet therapy</p> <p>Those patients with AF whose latest record of a CHADS2 score is > 1 should be offered anticoagulation therapy</p>

CHADS2, congestive heart failure, hypertension, age > 75 years, diabetes mellitus and prior stroke. Reproduced from Lawton *et al.*⁵² This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Sample

To gain a range of perspectives within practice teams, we aimed for a total sample comprising 30 GPs, 15 practice nurses and 15 practice managers. We invited staff from the 89 practices that shared data in WP2 to participate. Recruitment ran from September 2013 until June 2014.

Interview procedure

Each interview covered two indicators out of the eight under consideration. The topic guide drew on the TDF and was refined following piloting with three academic GPs⁵⁴ (see table 10 of Lawton *et al.*⁵²).

Data analysis

Interviews were audio-recorded and transcribed verbatim. We used NVivo 10 (QSR International, Warrington, UK) software to facilitate analysis. Initial coding was completed by the same three researchers who conducted the interviews (GL, JH and EI).

Data saturation was considered at the indicator level. Interviewers held regular debriefing discussions after interviews and reached consensus on new or redundant content across the indicators. Redundancy typically coincided with the estimated 15 interviews per indicator, ranging approximately between 12 and 14 interviews. However, for each indicator, we also sought further interviews with any less well-represented participant group to maximise the diversity of data.

Our framework analysis comprised familiarisation, identification of a framework, indexing, charting and mapping, and interpretation.⁵⁹ The coding framework was developed through an iterative process that incorporated the study aims, the TDF and the detailed reading of interview transcripts. The coding framework included code definitions to ensure consistency.

As part of familiarisation, researchers read through each transcript before coding and wrote a brief summary outlining key themes and findings. Pieces of text were coded according to the iterative coding framework. At this stage, TDF determinants (primary codes) were coded at a broad level. Common additional codes and categories were organised into secondary codes for TDF determinants. When additional codes and categories were added to the framework, we revisited coded transcripts and applied the revised coding framework. Early-stage, face-to-face meetings ensured that there was agreement in coding. To promote reliability, six transcripts were coded independently by each researcher and disagreements were resolved through discussion.

We completed two stages of analysis. We first assessed determinants for individual indicators by examining the data coded against the TDF domains. To prioritise TDF determinants for each indicator, we focused on the key thematic content (e.g. the extent to which the TDF determinant was discussed across participant groups) and barriers and enablers. We then identified meta-themes across multiple indicators, including the additional codes and categories generated to produce the analytical framework.

Results

We conducted 60 face-to-face interviews, with a ratio of 2 : 1 : 1 between GPs, practice managers and nurses, respectively, from a total of 31 general practices (Table 2). Interviews typically lasted around 30 minutes per indicator. Most participants were female (70%) and aged 40–49 years (38%). The mean number of years of experience in general practice was 14 years (range 1 to 33 years).

Theoretical domains

Professional role and identity and environmental context and resources featured prominently across all indicators, whereas the importance of other domains (e.g. beliefs about consequences, social influences and knowledge) varied across indicators. Table 4 from Lawton *et al.*⁵² describes all TDF domain content and specific barriers and enablers for each indicator and longer narrative accounts are also available in the published paper. We next focus on the meta-themes that emerged when synthesising data from multiple indicators.

TABLE 2 Allocation of interview topics by primary care staff (WP3a)

Recommendations	GP	Practice manager	Nurse	Total
Risky prescribing	8	3	4	15
Treatment targets in type 2 diabetes	7	4	4	15
BP targets in treated hypertension	7	4	4	15
Anticoagulation in AF	7	3	5	15
Total	29	14	17	60

Reproduced from Lawton *et al.*⁵² This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Meta-themes spanning multiple indicators

We identified five meta-themes that potentially represent general influences on evidence-based practice: (1) perceived nature of the job and norms of practice, (2) internal and external sources of support, (3) communication pathways and interaction, (4) meeting the needs of patients and (5) perceptions of indicators. Tables 5–9 in Lawton *et al.*⁵² illustrate interview excerpts.

Perceived nature of the job and norms of practice

When discussing the indicators and associated clinical behaviours, primary care professionals generally viewed the workload and burden associated with adherence as accepted and embedded components of general practice. Although professionals sometimes felt that indicators were imposed on consultations and that there was a limit as to what was achievable within a typical 10-minute consultation, they understood their roles in meeting QOF targets and recognised standards of practice. They further recognised that implementation could improve outcomes and reduce health-care costs in the longer term. Awareness of the indicators encouraged familiarity with required care processes and subsequent ingraining in everyday practice.

Although professionals described similar impacts of meeting the indicators, approaches to implementation differed between professional groups. Although GPs acted relatively autonomously and felt able to deviate from policies and procedures to tailor patient care, nurses preferred to follow policies and procedures, often justifying this approach by referring to risk and the threat of litigation. Some GPs felt that system prompts for implementing indicators disrupted consultations and sometimes directed their focus away from issues important to patients or patients' reasons for consulting. In contrast, many nurses said that they relied on templates and prompts to deliver appropriate care.

Internal and external sources of support

Professionals perceived both internal and external sources of support as critical to successful implementation. This often took the form of specific practice staff having specialised knowledge or lead roles for a clinical area. External support was provided by colleagues in secondary care or via network meetings with other practices. These sources provided trusted points of reference where professionals could seek the opinion of more knowledgeable colleagues and learn from others' experience. Other supports assisted implementation by prompting memory and regulating clinical behaviour. These were provided at the practice level by regular practice meetings and the development and use of internally developed prompts and templates, and at the wider organisational level via information technology and system infrastructure provided by Clinical Commissioning Groups (CCGs) and other bodies.

Communication pathways and interaction

Many professionals believed that effective interaction and information sharing were key to successful implementation. These required channels and skills to facilitate communication at three levels: between professionals and patients; between colleagues in a practice; and between primary and secondary care. Effective communication also depended on clear care pathways and respective professional roles; however, some professionals felt that there was scope for improving how communication systems provided support.

Meeting patient needs

Professionals evidently considered it important to take a holistic view of the patient when making decisions, irrespective of whether or not this resulted in deviating from recommended practice. This individualisation of patient care appeared to be driven by a strong sense of professional ethos and beliefs that it truly reflected quality of care and improved patient outcomes. Interviewees, particularly GPs, also acknowledged that patient priorities, preferences for treatment, and social and financial circumstances all influenced their practice and hence achievement of indicators. Although the latter factors were largely captured by the social influences TDF domain, other patient factors outside professional control influenced indicator achievement. These included patients' own knowledge around conditions, varying adherence to treatment and failures to attend pre-arranged consultations. Such influences appeared particularly relevant for indicators focused on outcomes (i.e. diabetes and BP control).

Perceptions of indicators

The content and structure of indicators and associated clinical practice recommendations were incompletely captured by the TDF. Although some recommendations, which were regarded as relatively clear and simple to follow, facilitated implementation, others were considered unnecessarily complex, lacking in clarity, or too lengthy – hindering their application in a time-pressured environment. There were also concerns about frequent revisions to recommendations and subsequent impacts on abilities to recall required procedures and processes, as well as perceived credibility of sources and recommendations.

Discussion

We identified a wide range of factors that can determine adherence to 'high-impact' indicators in primary care. Those related to social and professional roles and identity and environmental context and resources were prominent themes across all indicators, whereas the importance of other domains, for example beliefs about consequences, social influences and knowledge, varied across recommendations. We further identified five more general meta-themes important to primary care professionals in the implementation of all the indicators. Taken together, our findings suggested that it was feasible to develop implementation strategies for different evidence-based indicators that include both common features and content-specific adaptations.

Although some theoretical influences on adherence were shared across the four indicators, there were important variations; for example, environmental context and resources featured in discussions of all of the indicators. However, the specific belief contents varied considerably, with poor communication between primary and secondary care being a problem for prescribing anticoagulation for AF, whereas resource constraints, particularly the limited availability of ambulatory BP monitors, was identified for hypertension management. Social and professional roles and identity was also important across all indicators; some interviewees did not feel responsible for achieving some indicators. Other prominent determinants included beliefs about consequences, social influences, knowledge and memory, attention and decision processes, the latter being particularly relevant for prescribing decisions. Less evident domains included motivation, beliefs about capabilities, skills and emotion.

We identified five meta-themes from a synthesis of data across all four indicators, which broadly represent cultural, professional and system influences on evidence-based practice. Some of these might be amenable to change only at higher organisational levels (i.e. beyond the practice team), such as external sources of support and communication pathways, or even further upstream in the development

and dissemination of guidance, particularly perceptions of indicators.⁶⁰ Nevertheless, our findings underline the value of opportunities to share knowledge, expertise and support via local information technology systems for more efficient communication across care pathways.

Our interviewees consistently indicated the central role of patients for certain indicators, especially where outcomes partly or largely depend on patient behaviour. Many interviewees recognised the role of consultation and counselling skills in enabling patient behaviour change. First, patients influence professionals' decisions indirectly, sometimes via assumptions the latter make about the values and preferences of their patients. Second, the patient's own behaviour frequently featured as a barrier to indicator achievement; for example, BP control is more difficult to achieve if a patient drinks alcohol excessively or does not adhere to prescribed medication. Thus, the motivation and goals of both professionals and patients may need to be addressed simultaneously to optimise outcomes.⁶¹ Interventions that target both patients and professionals appear more likely to achieve glycaemic control in diabetes than those targeting either group in isolation.⁶²

Professionals often discussed general perceptions of guidelines and indicators. Many participants, particularly GPs, acknowledged that the value of guidelines was clear for the population but not for some patients, perhaps those with comorbidities or complex needs, whose adherence to recommendations could result in poorer outcomes. This perceived inflexibility has been reported in other studies of guideline adherence.^{63,64}

We highlight four limitations. First, the indicators we studied generally related to grouped behaviours or treatment goals (e.g. BP control in hypertension); thus, responses to questions rarely related to the enacting of a specific behaviour (e.g. taking a patient's BP during a consultation). The TDF is more typically proposed and used to investigate specific behaviours. Second, we actively encouraged participants to talk about each domain and analysed the data by looking for evidence that each domain was referenced in the language of participants. Although this may have prompted people to think about influences that might not come to mind (e.g. emotion), it made prioritising domains for intervention development difficult. Simply asking participants to talk about the factors that influence their behaviour may be a better technique for identifying key domains. Third, the TDF approach is based on the assumption that explanations of behaviour can be verbalised, that most individuals have the insight to do this and that these explanations resemble the actual influences on behaviour. Accepting the interview findings uncritically as 'the truth,' free of post hoc rationalisation, self-presentation bias and so forth, would be naive. Fourth, we acknowledge the significant influences of patients on health professionals' behaviour, and their role as actors in their own right. These both affect achievement of indicator targets; therefore, we may have identified further barriers and enablers had we also interviewed patients. Our findings suggest the potential value of interventions for selected indicators that target both patients and professionals.

Conclusion

We elicited a wide range of reported determinants of adherence to 'high-impact' indicators in primary care using the TDF. It was more difficult to pinpoint which determinants, if targeted by an implementation strategy, would maximise change. The meta-themes broadly underline the need to align the design of interventions targeting general practices with higher-level supports and broader contextual considerations. However, our findings suggested that it was feasible to develop interventions to promote the uptake of different evidence-based indicators that share common features while also including content-specific adaptations.

Work package 3b: developing an adaptable implementation package for indicators in primary care

Background

We next aimed to develop an implementation package that could be adapted to target each of four selected indicators. Accurate intervention descriptions can improve understanding of the effects of

interventions to change professional behaviour and hence guide their continuing optimisation. The Behaviour Change Taxonomy outlines 93 specific behaviour change techniques (BCTs) – observable, replicable and irreducible ‘active ingredients’ that offer a common language with which to describe intervention content.^{65,66} We planned to embed BCTs targeting determinants of adherence for each of four indicators within the adaptable implementation package.

Methods

We built our implementation package over five overlapping stages.

Stage 1: selecting delivery mechanisms

We selected delivery mechanisms typically available within primary care and of known effectiveness: audit and feedback (A&F),⁶⁷ educational outreach⁶⁸ and computerised prompts and reminders.^{69,70} We aimed to embed features associated with higher effectiveness (e.g. repeated feedback of audit data, requiring prescribers to select a reason for over-riding a computerised prompt).^{67,69}

Stage 2: identifying candidate behaviour change techniques

Team members (LG, RL, RM and RF) independently mapped BCTs (e.g. ‘feedback on behaviour’ or ‘action-planning’)⁶⁶ to theoretical domains⁵⁴ and resolved discrepancies by discussion. We thereby generated an inclusive list of ‘candidate’ change techniques.

Stage 3: prioritising determinants of behaviour

We convened a series of multidisciplinary panel meetings, one for each indicator. We invited 5–10 stakeholders with a range of perspectives and skills, including GPs, practice nurses, pharmacists, practice managers, quality improvement specialists and service commissioners. We presented them with emerging analyses from our earlier interviews (WP3a) with primary care professionals (frequency data and illustrative quotes for each determinant of achievement).⁵² After reviewing the range of determinants, stakeholders contextualised our findings and suggested additional professional or organisational determinants. The panel considered the feasibility and acceptability of candidate BCTs and intervention delivery mechanisms and their potential enhanced features, taking primary care context and resources into account. We took field notes of discussions. We convened our patient and public involvement (PPI) panel in parallel and followed similar methods. The research team communicated key messages from one panel to another and reviewed suggestions from both groups. Following the stakeholder panels, we further analysed interview findings to identify the most prominent determinants and high-level themes. We grouped determinants into four categories: core, prominent, less evident and not identified. Determinants considered core to all four QIs (i.e. consistently raised regardless of QI) included ‘social and professional role’, and ‘environmental context and resources.’ Those considered prominent (i.e. determinants that varied in importance) included ‘beliefs about consequences’, ‘social influences’, ‘knowledge’ and ‘memory, attention and decision processes.’ ‘Skills’, ‘beliefs about capabilities’ and ‘motivation and goals’ were less evident, whereas ‘emotion’ and ‘behavioural regulation’ were not identified.

Stage 4: designing intervention content

We drew on stages 1–3 to create a prototype outline for each delivery mechanism (feedback report, educational outreach session, and prompts and reminders). We did not develop computerised prompts for diabetes or BP control because they were already widely used to support QOF.¹⁰ Stakeholders also suggested patient-directed checklists to guide discussions around diabetes and BP control. We embedded candidate BCTs to target modifiable determinants of adherence. The prototype was adapted and tailored for each QI. We used the vocabulary and experiences expressed in interviews with health-care professionals and stakeholder panellists to tailor BCT content within delivery mechanisms. A graphic designer enhanced the intervention materials.

Stage 5: piloting and refining intervention content

We piloted each delivery mechanism for all QIs with five general practices involved in our earlier interview study (WP3a).⁵² A researcher (EI) directly observed the delivery of each educational outreach

session. She conducted brief, opportunistic semistructured interviews with practice staff (six GPs, two practice managers, and three practice nurses). Participants commented on the acceptability and feasibility of prototype feedback reports, patient-directed checklists and protocols for computerised prompts. We reviewed field notes and iteratively refined intervention content.

Results

We developed an implementation package adapted for each QI, fully described in Glidewell *et al.*⁷¹ (see *Appendix 1, Table 17*). Behaviour change technique categories with the potential to target one or more theoretical determinants were identified in a matrix (*Tables 3 and 4*). We identified 30 BCTs with potential to target determinants from our interview study and stakeholder panels (*Table 5*). We discarded those BCTs that could not be operationalised within our delivery mechanisms or existing primary care resources (see *Appendix 1, Table 18*).

The implementation package adapted for four quality indicators

Audit and feedback aimed to give comparative feedback on achievement, inform and prompt recall of clinical goals, highlight consequences of changing or not changing practice, suggest strategies for change and encourage goal-setting and reflection on progress towards goals. Reports included remotely gathered, individualised practice data and presented achievement for relevant trial indicators in graphical and numerical forms. Bar charts ranked practices by achievement and allowed comparisons with other (anonymised) trial practices in the same CCG and wider region, with accompanying text providing positive or encouraging feedback according to whether achievement had risen or fallen since the first report. Reports also contained brief, evidence-based clinical messages, responses to common queries (e.g. concerning data validity) and action-planning templates. Practices received reports quarterly in both electronic and paper form. Reports were accompanied by computerised search tools to identify relevant patients for review and significant event audit templates to support root cause analyses (risky prescribing; anticoagulation for AF).

Educational outreach aimed to build on feedback by facilitating individual and group reflection, discussing barriers to action, sharing models of good practice, enhancing motivation and action-planning. We trained pharmacists over 2 days to deliver sessions. The 30-minute sessions were designed to fit in with existing practice meetings and were offered to but not mandatory for intervention practices. We invited all staff involved in patient and practice management to attend. We identified a key clinical contact to support practice engagement. We offered a follow-up session to review progress and refine action plans, as well as 2 days of pharmacist support for patient identification and review.

Prompts and reminders aimed to reinforce clinical messages and indicator adherence. Computerised prompts for risky prescribing were triggered during consultations and repeat prescribing on the basis of an algorithm for patient age, diagnosis, drug and duration. A one-click justification (ignore, add or stop medication) was required before users could proceed.

The prompt for AF was not operationalised in time for trial evaluation. To avoid duplication with existing quality improvement systems, we did not develop computerised prompts for diabetes or BP control.⁷² We provided laminated reminders to convey key clinical information (e.g. management pathways) for BP control, anticoagulation for AF and risky prescribing. We provided pens and sticky notes containing key clinical messages to reinforce recommended practice. We developed patient-directed checklists to facilitate shared decision-making for BP control and diabetes control practices. However, we could not identify an efficient way to deliver these within routine consultations.

Identification of behaviour change techniques included within implementation packages

Each implementation package included at least 27 out of 30 potentially applicable BCTs (see *Table 5*), representing 15 of 16 BCT categories. Each package contained multiple unique instances of the

TABLE 3 Theoretical determinants of adherence to QIs and their prioritisation for intervention development (WP3b)

Layered identification of theoretical determinants	Capability				Opportunity		Motivation				
	Physical	Psychological			Social	Physical	Reflective			Automatic	Other
	Skills	Knowledge	Memory	Behavioural regulation	Social influences	Environmental context	Beliefs about capabilities	Beliefs about consequences	Social professional role	Emotion	Patient factors
1. Consensus panel of clinical and patient stakeholders	DC		DC								
2. Extended qualitative analysis of interview data		AF, RP, DC	AF RP	AF	BP, DC	All		All	All		All
Combined analysis	AF, BP, DC	All	AF, RP, DC	All	All	All	BP, RP, DC	All	All	AF, RP, DC	All
BP, blood pressure control; DC, diabetes control; RP, risky prescribing. Reproduced from Glidewell <i>et al.</i> ⁷¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.											

TABLE 4 Candidate BCTs mapped to determinants of adherence ordered by core, prominent and less evident determinants identified during interview study⁵² (WP3b)

Potential BCT categories ⁶⁶ ordered by likelihood of targeting core, prominent and less evident determinants	Core to all indicators		Prominent across indicators				Less evident		
	Environmental context	Social professional role	Knowledge	Memory	Social influences	Beliefs about consequences	Skills	Beliefs about capabilities	Motivation and goals
Social support	•	•			•			•	
Antecedents	•			•					
Comparison of behaviour		•	•		•		•	•	•
Feedback and monitoring		•		•		•		•	•
Identity		•							
Covert learning		•							
Comparison of outcomes			•		•	•			•
Natural consequences			•			•			•
Shaping knowledge			•				•		
Goals and planning				•				•	•
Repetition and substitution				•			•	•	
Associations				•					•
Regulation				•					
Reward and threat					•				•
Self-belief								•	
Scheduled consequences									•
Number of potentially relevant BCT categories	2	5	4	6	4	3	3	6	8

Reproduced from Glidewell *et al.*⁷¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

different BCTs. Four BCTs that were intended for inclusion ('identification of self as a role model' and 'verbal persuasion about capability' in educational outreach, 'discrepancy between current behaviour and goal' in feedback reports and 'anticipated regret' in feedback reports and educational outreach) could not be confirmed in a subsequent content check.

Extent of shared and unique behaviour change technique content across implementation packages

Twenty-three BCTs were shared across all QIs (see *Table 5*). Twenty-seven BCTs were identified in strategies targeting risky prescribing and BP control and 30 were identified in strategies targeting anticoagulation for AF and diabetes control. Seven BCTs were unique to implementation packages largely focused on changing processes of care (risky prescribing and anticoagulation for AF contained BCTs relating to 'goal-setting for behaviour' and 'monitoring of behaviour') and five BCTs were unique to packages targeting patient outcomes (BP control and diabetes control contained BCTs relating to 'goal-setting for outcomes' and 'monitoring for outcomes'). We did not operationalise 'goal-setting for behaviour' or 'monitoring of behaviours' for BP and diabetes control that focused on outcomes of behaviour.

Discussion

We aimed to provide a transparent account of intervention development and report sufficient detail for adoption, adaptation or evidence synthesis. We identified a large proportion of shared BCTs (at least 23 of 30 eligible BCTs) representing 15 of 16 BCT categories, suggesting that prioritised BCTs can be embedded and identified across delivery mechanisms adapted for different QIs.

We had to make trade-offs between what is theoretically desirable, clinically acceptable and operationally feasible in the context of delivery mechanisms and primary care resources. First, there were limitations in how we assessed and prioritised determinants of behaviour and subsequently linked them to BCTs. We used emerging and extended interview findings to inform intervention development. It was not possible within our research timelines to use the extended findings to inform adaptation of educational outreach or initial feedback reports. We may not have adequately operationalised BCTs to target core and prominent determinants ('social and professional role' and 'environmental context and resources') in the following categories: 'social support', 'antecedents', 'identity' and 'covert learning' to target the determinants. Second, BCTs from social cognition models and the TDF more generally focus on individual cognitions and may be insufficient to adequately target team, patient or organisational determinants. Third, although determinants of practice may be relevant only to countries with comparable primary care systems, methods to identify candidate BCTs and verify their presence are transferable.

We would have preferred to undertake more extensive piloting of the intervention as a whole and across all four targeted indicators. Our time to do this was limited by our decision to bring forward the trials' start date by 3 months so that the intervention period would coincide with the QOF year.

Conclusion

We have demonstrated the specification of BCT content for an adaptable implementation package. We identified variable numbers of BCTs but would not claim that 'more is better'; the ability to effectively target the most salient determinants is likely to be more important.

TABLE 5 Full description of intervention content by delivery mechanism and QI

Determinants of behaviour	BCTs verified by independent coder	BCT taxonomy code reference	Implementation package (see subsequent headings for variation in delivery mechanisms by QI)				A&F	
			Risky prescribing	Diabetes control	Anticoagulation	BP control	Risky prescribing	Diabetes control
'Environmental context', 'social and professional role' and 'social influences'	Social support	Social support unspecified (3.1)	•	•	•	•	•	•
		Social support practical (3.2)	•	•	•	•	•	•
'Environmental context' and 'memory'	Antecedents	Restructuring the physical environment (12.1)				•		
		Restructuring the social environment (12.2)		•				•
		Adding objects to the environment (12.5)						
'Social and professional role', 'knowledge' and 'social influences'	Comparison of behaviour	Social comparison (6.2)	•	•	•	•	•	•
		Information about others' approval (6.3)	•	•	•	•	•	•
'Social and professional role', 'memory' and 'beliefs about consequences'	Feedback and monitoring	Feedback on behaviour (2.2)	•	•	•	•	•	•
		Self-monitoring of behaviour (2.3)	•	•	•	•	•	•
		Self-monitoring of outcomes of behaviour (2.4)	•	•	•	•	•	•
		Feedback on outcomes of behaviour (2.7)		•		•		•
'Social and professional role'	Identity	Framing/reframing (13.2)			•			
	Covert learning	Vicarious consequences (16.3)	•	•	•	•	•	•
'Knowledge', 'social influences' and 'beliefs about consequences'	Comparison of outcomes	Credible source (9.1)	•	•	•		•	•
		Pros and cons (9.2)	•	•	•	•	•	
'Knowledge' and 'beliefs about consequences'	Natural consequences	Information about health consequences (5.1)	•	•	•		•	•
		Salience of consequences (5.2)	•	•	•	•	•	•
		Information about social/ environmental consequences (5.3)	•	•	•	•	•	•
'Knowledge'	Shaping knowledge	Instruction on how to perform the behaviour (4.1)	•	•	•	•	•	•
		Information about antecedents (4.2)	•	•	•	•	•	•
		Reattribution (4.3)	•	•	•	•	•	•

		Educational outreach				Computerised prompts and/or paper-based reminders			
Anticoagulation	BP control	Risky prescribing	Diabetes control	Anticoagulation	BP control	Risky prescribing	Anticoagulation	BP control	Diabetes control (not developed to prevent overlap with other QI initiatives)
•	1•	•	•	•	•				
•	•	•	•	•	•				
	•								
						•	•		
•	•	•	•	•	•				
•	•	•	•	•	•				
•	•								
•	•								
•	•								
	•								
•	•								
•	•								
•	•	•	•	•	•	•	•		
•		•	•	•	•	•	•		
•	•	•	•	•	•	•	•		
•	•								
•	•	•		•	•				
•	•	•	•	•	•				
•	•								
•	•								

TABLE 5 Full description of intervention content by delivery mechanism and QI (continued)

Determinants of behaviour	BCTs verified by independent coder	BCT taxonomy code reference	Implementation package (see subsequent headings for variation in delivery mechanisms by QI)				A&F	
			Risky prescribing	Diabetes control	Anticoagulation	BP control	Risky prescribing	Diabetes control
'Memory'	Goals and planning	Goal-setting behaviour (1.1)	•	•	•	•	•	•
		Problem solving (1.2)	•	•	•	•	•	•
		Goal-setting outcome (1.3)		•	•	•		•
		Action-planning (1.4)	•	•	•	•	•	•
		Review behavioural goals (1.5)	•	•	•	•	•	•
		Review outcome goals (1.7)		•	•	•		•
		Behavioural contract (1.8)	•		•		•	
		Commitment (1.9)	•		•		•	
	Repetition and substitution	Habit formation (8.3)		•				•
		Graded tasks (8.7)	•	•	•	•		•
	Associations	Prompts/cues (7.1)	•	•	•	•	•	•
	Regulation	Conserving mental resources (11.3)	•	•	•	•	•	•
'Social influences'	Reward and threat	Social reward (10.4)	•	•	•	•	•	•
'Beliefs about capabilities'	Self-belief	Focus on past success (15.3)	•	•	•	•	•	•
Total number of BCTs verified by independent coder			27	30	30	27	26	29
• = present.								

		Educational outreach				Computerised prompts and/or paper-based reminders				Diabetes control (not developed to prevent overlap with other QI initiatives)
Anticoagulation	BP control	Risky prescribing	Diabetes control	Anticoagulation	BP control	Risky prescribing	Anticoagulation	BP control		
•	•	•		•						
•	•	•	•	•	•					
•	•		•		•					
•	•	•	•	•	•					
•	•	•	•	•	•					
•	•		•	•	•					
•										
•										
•	•	•	•	•	•					
•	•	•	•	•	•	•	•		•	
	•					•	•			
•	•	•	•	•	•					
•	•									
29	28	16	16	17	17	6	6	1		

Work package 4a: a cluster-randomised evaluation of an adaptable implementation package targeting 'high-impact' evidence-based indicators in primary care

Background

Having developed an adaptable implementation package, tailored to target four high-impact indicators, we assessed its effects and cost-effectiveness in a cluster-randomised controlled evaluation. The published protocol is available as Willis *et al.*⁷³

Methods

Study design and setting

We conducted two parallel, cluster-randomised controlled trials (cRCTs) using balanced incomplete block designs. Cluster randomisation was essential as interventions were delivered at the general practice level (cluster). We maximised pragmatism in trial design and execution to ensure 'real-world' relevance.⁷⁴ Practices were recruited from West Yorkshire, England covering a socioeconomically diverse population of 2.2 million residents⁷⁵ that is broadly typical of national demographics, with the exception of higher deprivation levels.⁴⁴ Over 300 general practices are organised within 10 CCGs.

Each trial evaluated the effect of adapted implementation packages on adherence to two of four high-impact QIs: diabetes control and risky prescribing in trial 1 and BP control and anticoagulation in AF in trial 2. We selected the four indicators based on scope for improvement in practice (guided by WP2 findings) and potential population benefit through the achievement of recommended treatment goals for all of HbA_{1c}, BP and cholesterol in type 2 diabetes, avoidance of risky prescribing of NSAIDs and antiplatelet drugs,⁷⁶ anticoagulant prescribing for stroke prevention in AF,^{77,78} and achievement of recommended BP levels in patients with hypertension and others at high risk of cardiovascular events.⁷⁹ Our selection also took trial design into account. Within each trial, we assumed that any clinical effects of either implementation package would be independent of one another; thus, practices randomised to the implementation package for one indicator acted as control practices for the other implementation package and vice versa (Tables 6 and 7).

TABLE 6 Overview of cluster-randomised evaluation (WP4a): trial 1

Trial 1 ^a	Practices 1–40	Practices 41–80
Adapted implementation package for diabetes control	Intervention	Control
Adapted implementation package for risky prescribing	Control	Intervention
a A further 34 general practices were randomised as non-intervention controls.		

TABLE 7 Overview of cluster-randomised evaluation (WP4a): trial 2

Trial 2 ^a	Practices 81–112	Practices 113–144
Adapted implementation package for BP control	Intervention	Control
Adapted implementation package for anticoagulation in AF	Control	Intervention
a A further 34 general practices were randomised as non-intervention controls.		

General practices were eligible if they used SystmOne, the computerised clinical system used by approximately two-thirds of West Yorkshire practices [The Phoenix Partnership, URL: www.tpp-uk.com (accessed 16 September 2019)]. We excluded practices involved in intervention development and piloting (i.e. WPs 2 and 3).

We used an opt-out approach to practice recruitment to facilitate participation and enhance generalisability.⁸⁰ We invited eligible practices to participate via recorded post and e-mail, with reminders at 2 weeks to non-responding practices. We included those which had not actively declined by 4 weeks. The use of opt-out recruitment avoided the biases associated with opt-in approaches and enhanced sample representativeness.^{44,48} We obtained anonymised patient-level data⁸¹ through data extracts from SystmOne.

Procedures

We adapted the implementation package for each QI comprising a combination of A&F reports, educational outreach visits and computerised prompts with embedded BCTs, as detailed in WP3b.⁶⁶ A dedicated administrator contacted practices to offer to co-ordinate the educational outreach visits, which were also promoted in feedback reports and e-mail communications to all practices.

Trial end points

Table 8 lists all outcomes. In brief, primary outcomes at 11 months post randomisation were achievement of all recommended target levels of HbA_{1c}, BP and cholesterol in patients with type 2 diabetes; a composite indicator of risky prescribing; achievement of recommended BP targets for specific patient groups; and anticoagulation prescribing in eligible patients with AF.

We modified three of the indicators representing the primary outcomes from those we developed and assessed in earlier WPs:

- Risky prescribing. We modified 'Prescribing of warfarin and a traditional oral NSAID without coprescription of a gastroprotective drug' to 'Prescribing of warfarin and a traditional oral NSAID.' Combined use of warfarin and NSAIDs is contraindicated and it would not make clinical sense to add gastroprotection in this situation.
- BP control. We broadened the indicator that had focused on targets in treated hypertension to encompass BP control across a range of conditions associated with higher cardiovascular risk (e.g. CKD) as well as patients with a cardiovascular disease risk of $\geq 20\%$. We judged that this would potentially permit a greater population impact and assumed that the determinants of practice were unlikely to change markedly.
- Anticoagulation for AF. The indicator 'Those patients with AF whose latest record of a congestive heart failure, hypertension, age > 75 years, diabetes mellitus and prior stroke (CHADS₂) score is greater than one should be offered anti-coagulation therapy' was superseded by one using the congestive heart failure, hypertension, age > 75 years, diabetes mellitus, stroke and vascular disease (CHA₂DS₂-VASc) score. We also dropped the specification that 'Warfarin should be administered as the most effective thromboprophylactic agent' to reflect emerging guidance allowing prescribing of non-vitamin K antagonist oral anticoagulants (NOACs) as an alternative.⁸²

Secondary outcomes included individual indicators within the composite primary outcomes, processes of care and continuous clinical outcomes such as HbA_{1c}, BP and cholesterol.

Sample size

We used WP2 data to inform trial sample sizes. Mean cluster size (number of eligible patients per practice), cluster size coefficient of variation, intracluster correlation coefficient and mean achievement rates were calculated for each primary indicator (Table 9).

TABLE 8 Primary and secondary outcomes in cluster-randomised evaluation of implementation package (WP4a)

Diabetes control	Risky prescribing	BP control	Anticoagulation in AF
Primary outcomes			
<p>The proportion of patients with type 2 diabetes achieving all three of the following treatment targets:</p> <ol style="list-style-type: none"> 1. BP of < 140/80 mmHg (or 130/80 mmHg if there is kidney, eye or cerebrovascular damage) 2. HbA_{1c} value of ≤ 59 mmol/mol 3. Total serum cholesterol level of ≤ 5.0 mmol/l 	<p>The proportion of patients achieving at least one of the nine following indicators of high-risk NSAID and antiplatelet prescribing:</p> <ol style="list-style-type: none"> 1. Prescribing a traditional oral NSAID or low-dose aspirin in patients with a history of peptic ulceration without coprescription of gastroprotection 2. Prescribing a traditional oral NSAID in patients aged ≥ 75 years without coprescription of gastroprotection 3. Prescribing a traditional oral NSAID and aspirin in patients aged ≥ 65 years without coprescription of gastroprotection 4. Prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of gastroprotection 5. Prescribing of warfarin and a traditional oral NSAID 6. Prescribing of warfarin and low-dose aspirin or clopidogrel without coprescription of gastroprotection 7. Prescribing an oral NSAID in patients with heart failure 8. Prescribing an oral NSAID in patients prescribed both a diuretic and an angiotensin-converting-enzyme inhibitor (ACE-I) or angiotensin receptor blocker 9. Prescribing an oral NSAID in patients with CKD 	<p>The proportion of patients achieving at least one of the eight following recommended targets for satisfactorily controlled BP:</p> <ol style="list-style-type: none"> 1. BP of < 140/90 mmHg in patients aged < 80 years with hypertension 2. BP of < 150/90 mmHg in patients aged ≥ 80 years with hypertension 3. BP of < 140/80 mmHg in patients aged < 80 years with diabetes, or < 130/80 mmHg if there are complications of diabetes 4. BP of < 130/80 mmHg in patients aged < 80 years with CKD and proteinuria 5. BP of < 140/90 mmHg in patients aged < 80 years with coronary heart disease 6. BP of < 140/90 mmHg in patients aged < 80 years with peripheral arterial disease 7. BP of < 140/90 mmHg in patients aged < 80 years with a history of stroke or transient ischaemic attack 8. BP of < 140/90 mmHg in patients aged < 80 years with a cardiovascular disease risk of ≥ 20% <p>Note: patients should appear only once and, where any appear in more than one group with differing BP targets, the lower target should take precedence</p>	<p>The proportion of patients prescribed anticoagulation therapy in the following groups:</p> <ol style="list-style-type: none"> 1. Men with AF and a CHA₂DS₂-VASc score of 1 prescribed anticoagulation therapy 2. All people with AF and a CHA₂DS₂-VASc score of ≥ 2 prescribed anticoagulation therapy
Secondary outcomes			
<ol style="list-style-type: none"> 1. The proportion of patients with type 2 diabetes achieving each of the individual treatment targets that make up the primary outcome 	<ol style="list-style-type: none"> 1. The proportion of patients achieving the individual indicators that make up the primary outcome 	<ol style="list-style-type: none"> 1. The proportion of patients achieving the individual targets that make up the primary outcome 	<ol style="list-style-type: none"> 1. The proportion of patients achieving the indicators that make up the primary outcome

TABLE 8 Primary and secondary outcomes in cluster-randomised evaluation of implementation package (WP4a) (continued)

Diabetes control	Risky prescribing	BP control	Anticoagulation in AF
2. The proportion of patients with type 2 diabetes receiving each of the following nine processes of care in the previous 12 months: <ol style="list-style-type: none"> BP measured HbA_{1c} levels measured Full lipid profile performed Micro-albuminuria testing performed Estimated glomerular filtration rate/serum creatinine testing performed Foot care review Eye screening BMI recorded Smoking status recorded 	2. The proportion of patients achieving at least one of indicators 1–6 (gastrointestinal composite) 3. The proportion of patients achieving at least one of indicators 8 and 9 (renal composite)	2. Mean patient levels of: <ol style="list-style-type: none"> Systolic BP Diastolic BP 3. The proportion of patients with hypertension or at high risk of cardiovascular events with measured BP in the previous 12 months	2. The proportion of patients eligible for the two indicators who have contraindication for anticoagulation, i.e.: <ol style="list-style-type: none"> Men with AF and a CHA₂DS₂-VASc score of 1 with contraindication for anticoagulation All people with AF and a CHA₂DS₂-VASc score of 2 or above with contraindication for anticoagulation 3. The combined proportions of patients achieving the two indicators related to contraindication for anticoagulation as defined above in 2(a) and 2(b)
3. The proportion of patients with type 2 diabetes receiving all of the above processes of care			
4. The proportion of patients with type 2 diabetes receiving all of the processes of care as defined in above with the exception of (g) eye screening			
5. Mean patient levels of: <ol style="list-style-type: none"> Systolic BP Diastolic BP HbA_{1c} Total serum cholesterol 			

BMI, body mass index; CHA₂DS₂-VASc, congestive heart failure, hypertension, age > 75 years, diabetes mellitus, stroke and vascular disease.

TABLE 9 Key sample size assumptions for cluster-randomised evaluation of implementation package (WP4a)

	Diabetes control	Risky prescribing	BP control	Anticoagulation in AF
Mean cluster size (number of eligible patients per practice)	280	420	800	55
Coefficient of variation of cluster size	0.60	0.65	0.67	0.79
Intraclass correlation coefficient	0.06	0.03	0.06	0.06
Mean achievement (%)	43.0	89.0	72.0	60.0

Median effect sizes on processes and outcomes of care for a range of guideline implementation studies are around 4–9%.^{83,84} Given that we were evaluating enhanced, multifaceted interventions and targeting indicators with scope for improvement, we judged an absolute difference of 15% for diabetes control, BP control and AF outcomes as realistic and relevant from a population perspective. Control group achievement rates in risky prescribing were higher and considering a potential ceiling effect, we considered a 5% difference as realistic and clinically relevant. To achieve 90% power, and allowing for an alpha error rate of 2.5% (to adjust for two outcome comparisons in each trial) and a 10% drop-out rate, we required 40 practices per arm in trial 1 (diabetes and risky prescribing) and 32 practices per arm in trial 2 (BP control and anticoagulation for AF). We therefore aimed to recruit 144 practices.

The opt-out approach resulted in 178 practices being recruited, allowing randomisation to a fifth arm: a non-intervention control group to further assess Hawthorne effects.

Randomisation

A two-stage minimisation process (incorporating a random element) was undertaken centrally by the trial statistician. First, practices were stratified by CCG and list size, and randomised to trial 1, trial 2 or non-intervention group (80 : 64 : 34). Practices from one CCG involved in a concurrent initiative targeting anticoagulation in AF were ineligible for trial 2 and allocated to either trial 1 or non-intervention. Second, practices within each trial were randomised (1 : 1) to individual trial arms, minimised by CCG, list size and pre-intervention adherence to the two targeted indicators relevant to that trial. General practices and trial personnel involved in intervention delivery were, of necessity, made aware of allocation but collection of outcomes for the primary end points was blind.

Data collection

We gathered anonymised outcome data remotely from general practices, mostly using data collected for QOF between 1 April 2015 and 31 March 2016.

We obtained data on practice characteristics from publicly available sources (Health Education England, Health and Social Care Information Centre)^{45,85,86} including practice list size (number of registered patients); number of GP partners and salaried GPs; practice training status; practice-level IMD, ethnic profile of practice register, achievement of QOF indicators (2014–15); patient satisfaction (proportion who would recommend the practice to others); patient-rated practice accessibility (proportion able to speak with GP or nurse within 48 hours of approach); and practice prescribing costs. Patient characteristics were extracted along with QOF achievement data and included age, sex and comorbidity (number of QOF registers on which patient appeared).

Statistical analysis

We undertook all data summaries and analyses on the intention-to-treat population, defined as all patients registered at randomised practices, regardless of intervention uptake or loss from the trial. Statistical testing was completed at a two-sided 2.5% significance level; effect sizes and 97.5% confidence intervals (CIs) were reported in each case. All comparisons were conducted using the within-trial randomised controls. Analyses utilising the non-intervention controls are reported separately.

Data completeness for analyses depended on the completeness of SystmOne medical records and could not be assessed within the trial data set. Missing Read codes could lead to patients being incorrectly included in or excluded from eligible populations. Similarly, indicator achievement may have been incorrectly specified as a result of missing data. For the primary and secondary analyses, we assumed that data were missing at random.

We compared primary outcomes between intervention and control practices using two-level binary logistic regression models, with patients (level 1) nested within registered practices (level 2). We adjusted analyses for patient-level covariates (gender, baseline age) and practice-level covariates

(baseline practice list size, CCG, pre-intervention achievement against primary outcomes, overall QOF score 2014–15 and baseline proportion of patients with 0–3 comorbidities).

We analysed binary secondary outcomes for individual indicators within the composite primary outcomes and recorded processes of care using similar multilevel logistic models. We analysed continuous intermediate clinical outcomes using two-level linear models. We adjusted each model for the same covariates specified for the primary outcome analysis. Intervention effects on related wider practice behaviour (i.e. other indicators not targeted by the interventions) were assessed by QOF 2015–16 indicators mapped to trial outcomes. We assessed unintended effects on quality of care via a series of non-trial related QOF indicators in coronary heart disease, mental health, smoking and asthma (see table 1 of Willis *et al.*⁷³). We used practice-level achievement of each QOF indicator as the outcome in a linear model, adjusted for practice-level covariates (baseline practice list size, CCG, pre-intervention achievement against primary outcomes and overall QOF score 2014–15). Age and gender are patient-level covariates and were, therefore, not included in the analyses of QOF indicators.

We undertook appropriate regression diagnostics for binary and continuous outcomes to check the validity of the statistical modelling. We applied log transformations to continuous outcomes in cases where residual normality assumptions appeared to be violated.

As part of our process evaluation, we collected data to explore the fidelity of delivery, receipt, use and sustainability in both trial and process evaluation practices (described under WP5).

All analyses were planned prior to final data extraction and no interim analyses were planned or conducted. Statistical analyses were conducted in SAS® (SAS Institute Inc., Cary, NC, USA) software version 9.4 (SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration).

Results

Screening and recruitment

Between February 2015 and April 2015, we screened 278 SystmOne general practices across 10 West Yorkshire CCGs (*Figure 1*). A total of 243 (87.41%) were eligible and invited to participate; 56 (23.05%) of these practices opted out, largely because of workload pressures, and nine were excluded for other reasons (see *Report Supplementary Material 1, Tables 1–5*). Typically practices opted out within 2 weeks of receiving the invitation letter (mean 12.51 days, SD 11.47 days); thus, 178 (73.25%) of eligible practices were randomised. Practice characteristics were broadly similar between those not randomised and those randomised (see *Appendix 1, Table 19*).

Forty practices were allocated to each of the diabetes control and risky prescribing arms in trial 1, 32 practices were allocated to each of the BP control and anticoagulation in AF arms in trial 2. A further 34 practices were allocated to a non-intervention control arm (see *Report Supplementary Material 1, Tables 6 and 7*). None of the randomised practices withdrew from any aspect of the trial. One practice in the risky prescribing arm merged with a non- Action to Support Practices Implementing Research Evidence (ASPIRE) practice following the third feedback report at 9 months; no outcome data were available after 9 months. One practice in the BP control arm closed during the study because of GP retirement. This practice received two of the four planned feedback reports; outcome data were unavailable beyond 6 months' follow-up. Two ASPIRE practices merged during the trial, one from diabetes control and one from risky prescribing; all patients from both practices became registered at the diabetes control practice. Both practices received all four feedback reports and final outcome data were available for both practices. One further ASPIRE practice from risky prescribing merged with a non-ASPIRE practice. This practice received three of the four planned feedback reports; outcome data were unavailable beyond 9 months' follow-up.

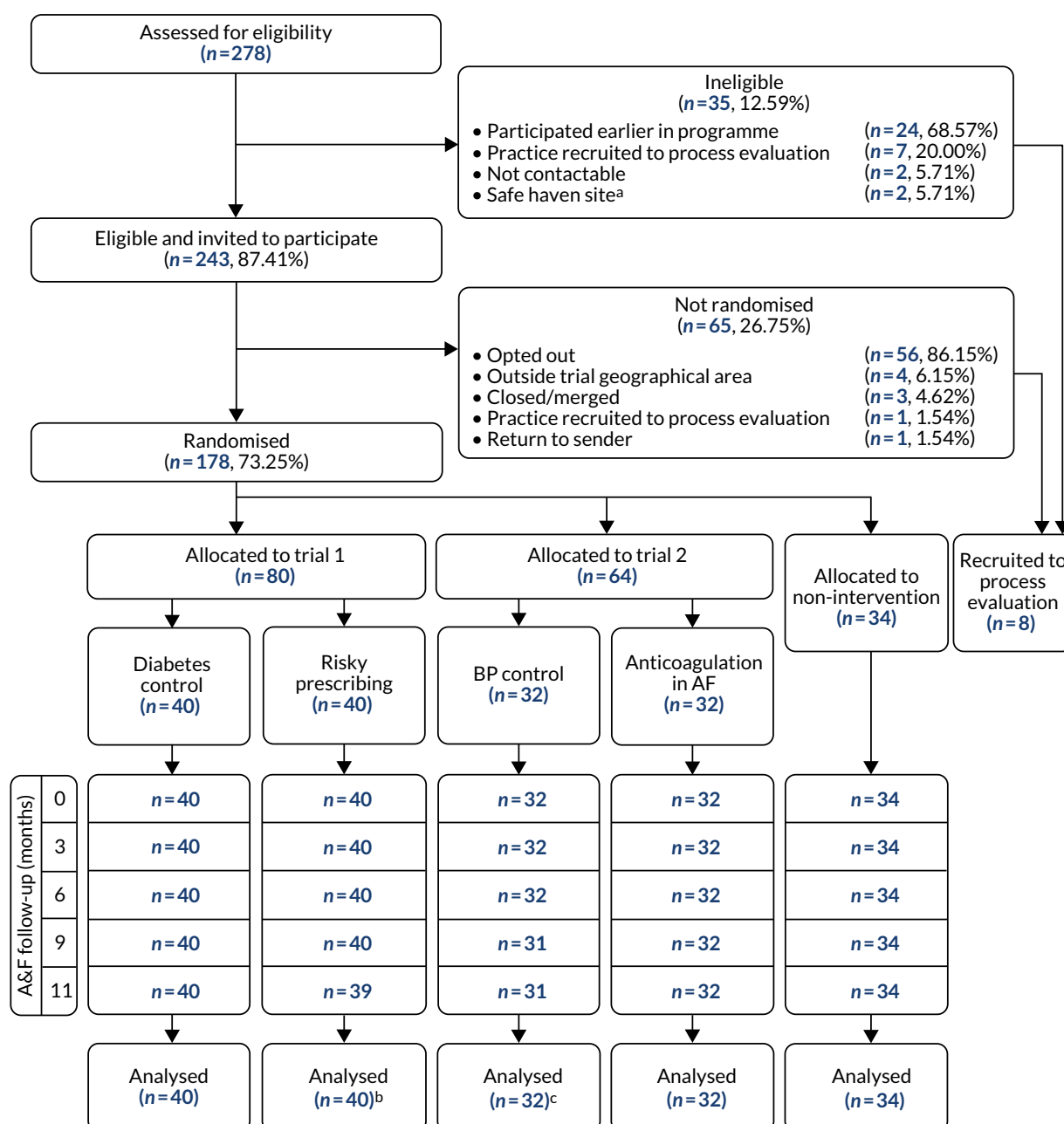


FIGURE 1 The CONSORT flow diagram for cluster-randomised evaluation (WP4a). a, Safe haven practices are for those patients who have demonstrated violent tendencies and have been removed from usual general practice or for ex-offenders; b, one practice in the risky prescribing arm merged with a non-Action to Support Practices Implementing Research Evidence (ASPIRE) practice in advance of the fourth feedback report; however, as they received the first three feedback reports some outcome data are available and they are included in the final analyses; c, one practice in the BP control arm closed in advance of the third feedback report; however, as it received the first two feedback reports some outcome data are available and these are included in the final analyses. Adapted from Willis *et al.*⁹⁷ © 2020 Willis *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Baseline practice characteristics were well balanced by trial and indicator (see *Appendix 1, Tables 19 and 20*, and *Report Supplementary Material 1, Table 8*): across all 178 practices, the mean list size was 7249.76 patients (SD 4306.72 patients); the mean IMD was 30.85 (SD 13.70, falling within the top quarter of social deprivation); and mean pre-intervention adherence was 33.29% (SD 7.08%) for diabetes control, 7.91% (SD 4.07%) for risky prescribing, 65.87% (SD 6.67%) for BP control and 66.20% (SD 11.82%) for anticoagulation in AF.

Baseline characteristics for patients within the populations relevant to the four indicators, as well as all patients within the practices, were broadly similar between the trial arms (see *Appendix 1, Table 21*, and *Report Supplementary Material 1, Tables 9–12*). The data obtained relating to polypharmacy (number of repeat prescriptions per patient) were found to be incomplete and, hence, were not used in any analyses.

Primary outcomes

The primary outcome results are summarised in *Table 10*, with detailed summaries in *Appendix 1, Table 22*. In trial 1, achievement of diabetes control at 11 months for practices assigned to this implementation package was 24.19% compared with 23.74% for practices assigned to the control arm (risky prescribing); an adjusted OR of 1.03 (97.5% CI 0.89 to 1.18), thus providing no evidence of an intervention effect.

There was evidence of a significant intervention effect for the risky prescribing implementation package. The proportion of patients with a record of risky prescribing at 11 months in practices assigned to this implementation package was 4.94% compared with 5.99% for control practices (diabetes control); an adjusted OR of 0.82 (97.5% CI 0.67 to 0.99). Thus, the odds of risky prescribing for a patient in an intervention practice were 18.5% lower (i.e. better) than for a patient with the same characteristics in a control practice.

In trial 2, there was no evidence of a significant effect on primary outcomes for either implementation package. Achievement of BP control at 11 months for practices assigned to this implementation package was 53.62% compared with 52.32% for practices assigned to the control arm (anticoagulation in AF); an adjusted OR of 1.05 (97.5% CI 0.96 to 1.16). Achievement of anticoagulation in AF at 11 months for practices assigned to this implementation package was 73.20% compared with 75.18% for control practices (BP control); an adjusted OR of 0.90 (97.5% CI 0.75 to 1.09).

TABLE 10 Primary outcome achievement: baseline rates, adjusted outcome rates and adjusted odds (WP4a)

Primary outcomes	Baseline achievement ^a (%)	Adjusted outcome achievement ^b (%)	Adjusted OR (97.5% CI)	NNT ^c	p-value
Trial 1: diabetes control					
Intervention	33.72	24.19	1.03 (0.89 to 1.18)	223	0.697
Control (risky prescribing)	34.43	23.74	1 [reference]		
Trial 1: risky prescribing					
Intervention	7.20	4.94	0.82 (0.67 to 0.99)	95	0.017
Control (diabetes control)	7.43	5.99	1 [reference]		
Trial 2: BP control					
Intervention	66.66	53.62	1.05 (0.96 to 1.16)	77	0.215
Control (anticoagulation in AF)	65.50	52.32	1 [reference]		
Trial 2: anticoagulation in AF					
Intervention	66.42	73.20	0.90 (0.75 to 1.09)	-51	0.2141
Control (BP control)	67.54	75.18	1 [reference]		
NNT, number needed to treat.					
a Calculation of achievement for diabetes control and BP control at baseline uses 'any' BP measurements taken in the previous 12 months.					
b Calculation of achievement for diabetes control and BP control at outcome uses 'the most recent' BP measurements taken.					
c The NNT has been rounded up (away from zero) to an integer value.					

Secondary binary outcomes

In trial 1, secondary analyses of individual indicators forming the composite primary outcomes and recorded processes of care showed no statistically significant effects relating to the diabetes control implementation package (see *Report Supplementary Material 1, Tables 13–26*).

The risky prescribing implementation package resulted in significantly lower (i.e. better) levels for one individual indicator, 'Patients aged 65 years or over prescribed aspirin and clopidogrel without coprescription of gastroprotection'. Prescribing levels were 25.32% in implementation package practices compared with 35.20% in control practices: an adjusted OR of 0.62 (97.5% CI 0.39 to 0.99). Thus, the odds of risky prescribing for a patient in an intervention practice was 37.59% lower (i.e. better) than for a patient with the same characteristics in a control practice (see *Report Supplementary Material 1, Table 30*). We identified no other statistically significant effects relating to other risky prescribing indicators (see *Report Supplementary Material 1, Tables 27–37*). In trial 2, we found no evidence of significant intervention effects on any individual indicators or processes of care (see *Report Supplementary Material 1, Tables 38–51*).

Secondary continuous outcomes

There were no significant intervention effects in the secondary analyses of the continuous intermediate clinical outcomes: systolic and diastolic BP (in both the diabetes control and the BP control populations); HbA_{1c} and total serum cholesterol (diabetes control only; see *Report Supplementary Material 1, Tables 52–57*). The conclusions for total serum cholesterol and HbA_{1c} were robust to log-transformation of the outcomes in a sensitivity analysis conducted following model diagnostic checks.

Secondary analysis of Quality and Outcomes Framework indicators

We explored intervention effects for QOF indicators to assess whether or not there were any benefits in the achievements of QOF indicator groups relevant to any of the implementation packages and whether or not there were any unintended harms of reduced achievement in QOF indicator groups not relevant to any of the packages (summary statistics; see *Report Supplementary Material 1, Tables 58–61*). Statistical modelling was not possible for all indicators because of violation of modelling assumptions, particularly where large proportions of practices attained 100% achievement for an indicator. For those indicators where modelling was appropriate, we found no statistically significant intervention effects (see *Report Supplementary Material 1, Tables 62–75*).

Intervention fidelity

Full analyses of intervention fidelity are presented under the process evaluation (WP5). In brief, all practices received A&F reports as intended (with the exception of those that closed or merged). Outreach visits were delivered to 67 (46.53%) practices, with 8 (5.56%) receiving two visits. There was similar uptake between trial arms (see *Appendix 1, Tables 26 and 27*). The offer of visits was declined by 77 practices, 52 (67.53%) of which did not give a reason (see *Report Supplementary Material 1, Table 76*). One hundred and twenty-six (87.50%) practices activated access to searches and computerised prompts (risky prescribing practices only; see *Report Supplementary Material 1, Tables 102–103*).

Discussion

An adaptable, multifaceted implementation package improved clinical care for one of four high-impact indicators in general practices serving relatively socially deprived populations. The odds of risky prescribing for a patient in an intervention practice was 18.50% lower than for a patient with the same characteristics in a control practice, with a population impact likely to translate into reduced mortality, morbidity and unplanned emergency admissions. There was no effect on diabetes control, BP control or anticoagulation in AF. The varying effects are likely to be attributable, in part, to differences in targeted behaviours, suggesting that in broad terms our adapted 'one-size-fits-all' approach was not universally effective.

It initially appeared as if one risky prescribing subindicator associated with a statistically significant effect might have driven the composite indicator. For 'patients aged 65 years or over prescribed aspirin and clopidogrel without coprescription of gastroprotection', the number of eligible patients was relatively small (334 in the intervention arm and 405 in the control arm) compared with the much larger overall eligible number of patients for the composite primary outcome (18,313 in the intervention arm and 18,131 in the control arm). However, the risky prescribing intervention had cumulative effects across most of the nine subindicators.

Our work is best contextualised within the wider literature related to key features of our implementation package, namely A&F, educational outreach visits, and prompts and reminders. Ivers *et al.*⁶⁷ found that A&F was more effective for targeting prescribing behaviours versus other outcomes related to diabetes and cardiovascular disease management. Similarly, reviews of educational outreach visits and computerised prompts suggested more reliable effects in changing simpler as opposed to more complex outcomes.^{68,70} However, these reviews include potentially confounded indirect comparisons between studies; our programme of trials is novel in allowing direct comparisons of effects between different targeted clinical behaviours, notwithstanding some tailoring of content according to each high-impact indicator. Our findings, therefore, suggest that the nature of the clinical behaviours targeted can be as important as the type of intervention in predicting the effects of implementation strategies.

Our findings are consistent with those of Dreischulte *et al.*,⁸⁸ who first demonstrated that an intervention including feedback reduced risky prescribing. However, where changes in patient as well as professional behaviour are important in achieving treatment goals, such as for long-term conditions like type 2 diabetes, interventions targeting both patients and professionals are more likely to be effective than those targeting professionals alone;⁸⁹ our implementation package mainly focused on professionals.

Our robustly designed trials used balanced incomplete block designs to permit comparison of intervention effects while minimising any potential Hawthorne effects.⁹⁰ Furthermore, the trials were highly pragmatic in three ways. First, opt-out recruitment is likely to have ensured that participating practices were representative of the wider population.^{44,48} Second, we used minimally intrusive data collection. Third, for intervention delivery all practices received, but were not obliged to read, feedback reports; outreach visits were optional.⁷⁴ Hence, the implementation packages were tested under 'real-world' conditions, increasing confidence in wider applicability to routine general practice settings.

Our evaluation had five main limitations. First, the use of routinely collected data may have compromised the precision of our outcomes and hence ability to demonstrate effects. Second, the multifaceted nature of the implementation package precludes any attempts to quantify the effects of individual intervention components. Third, educational outreach visits were delivered by facilitators who were not allocated to specific arms of the trial, thereby risking contamination between arms. We had instructed facilitators to focus only on delivering the implementation package to which each practice was assigned. Fourth, our follow-up period of 11 months may have been too short to ensure detection of changes in clinical outcomes, such as those related to diabetes or BP control, especially if a number of general practices only received educational outreach visits later in this period. However, other trials have demonstrated changed clinical outcomes within similar durations of follow-up⁸⁹ and we also detected no improvements in processes of care for diabetes and BP control. Fifth, our composite end point for diabetes control requiring achievement of treatment goals for all of HbA_{1c}, BP and cholesterol may have been too demanding. Nevertheless, we observed no improvements in any of these three contributing indicators separately and this end point was considered fair, if challenging, by our clinical and patient advisors.

Our work highlights three methodological controversies and challenges. First, the effect of our implementation package on risky prescribing was modest but important at a population level. Forgoing randomised designs, as some have suggested,^{91,92} would have reduced confidence in the validity of our findings, and risked false positive conclusions.⁹³ Second, although we see our pragmatic design as a strength, we could have applied a more explanatory approach and made full engagement with our

implementation package a condition of trial participation. However, such mandating is seldom possible or even desirable in quality improvement programmes that depend on professional consensus, particularly as it may encourage 'gaming' behaviours to achieve goals while circumventing real action. Similarly, opt-out recruitment may have diluted the contribution of self-selected and more enthusiastic practices, as well as reduced administrative burden for participants; however, those responsible for leading quality improvement initiatives often wish to specifically include less enthusiastic or poorer performing practices. Such an approach appears no longer feasible under current research governance arrangements, which require explicit permission from general practices to participate in research (a de facto opt-in approach).⁹⁴ Third, a critical challenge prior to pragmatic evaluations is to develop interventions that are sufficiently feasible and durable to withstand the relatively harsh environments of busy clinical practice. Although we largely followed the UK Medical Research Council framework for the development and evaluation of complex interventions,⁹⁵ practice engagement with our implementation package was highly variable. We would now recommend more intensive, iterative cycles of testing and refining interventions prior to scaling up within definitive evaluations.

Conclusion

We have demonstrated the effectiveness of a multifaceted implementation package in reducing risky prescribing. Our findings are directly applicable to general practice given the highly pragmatic nature of our evaluation. Interventions involving A&F have also been shown to reduce other undesired prescribing behaviours, namely of antibiotics,⁹⁶ and offer a means to address other urgent priorities, such as rising opioid prescribing.⁹⁷ However, there are still major challenges in addressing other high-impact indicators that are likely to require better targeted interventions that can be sustainably embedded within general practice.

Work package 4b: economic evaluation

Introduction

We conducted economic evaluations alongside the ASPIRE trial analyses. We used the trials data to parameterise decision-analytic models as the trials did not collect patient-level data. We did not expect any longer-term intervention benefits to be realised beyond the trials period and thus modelled any such impacts.

Given resource and time constraints, we were unable to evaluate all four trial indicators and prioritised the economic evaluations based on ASPIRE team and steering committee advice while accounting for the availability of 'off-the-shelf' decision models. There was a consensus that the risky prescribing indicator was the first choice for evaluation, followed by BP control; hence, we evaluated these two indicators and built associated de novo decision models. As a model was freely available for type 2 diabetes (UK Prospective Diabetes Study version 2),⁹⁸ we also evaluated this indicator.

Methods

General approach and model choice

We treated the implementation packages as separate interventions, and hence did not aggregate costs and effects. We aggregated subindicators for relevant composite indicators. Since we were evaluating the implementation of NICE recommendations, we aimed to mirror evaluative tools used in the evidence syntheses underpinning NICE guidelines. This was not possible for risky prescribing since this was a new indicator although we relied heavily on previous modelling for NSAIDs.

Where possible we followed the reference case set out by NICE.⁹⁹ As such, we reported cost per incremental quality-adjusted life-year (QALY) from a health and personal social service provider perspective. We adopted a lifetime perspective and discounted costs and benefits post 1 year at the NICE preferred rate of 3.5% per annum. Where we developed or adapted models, we conducted targeted searches to

identify parameter values. We developed and parameterised the models iteratively with ASPIRE team clinicians (RF, SA and DP). The decision models allowed the generation of cost per QALY estimates by translating trial effectiveness results on implementation and clinical outcomes into event risks and expected costs, survival and quality-of-life impact. We estimated costs where necessary using standard sources including *NHS Reference Costs*¹⁰⁰ and the Personal Social Services Research Unit report.¹⁰¹ We converted relevant historical or non-UK costs to 2017 UK prices using a health-related inflation tool.¹⁰²

We report incremental cost-effectiveness ratios (ICERs) per QALY gain in each evaluation. To aggregate cost-effectiveness within indicators (for risky prescribing and BP control), we generated weighted averages of costs and benefits based on the proportion of a general practice list eligible for each indicator. ICERs below the NICE preferred threshold of £20,000–30,000 cost per QALY were taken to indicate cost-effectiveness. We ran deterministic sensitivity analyses to test the impact of changes in parameter values and assumptions. Where possible, we also ran probabilistic sensitivity analyses to test the overall impact of parameter uncertainty on model results. We present these results as cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs).¹⁰³

Intervention costing

The ASPIRE implementation package costs comprised fixed and variable components that were all expressed at practice level since all reported outcomes, namely QALY gain and health-care savings, were computed for the average practice list size of 7130 patients. The fixed cost consisted of preparing, delivering and receiving the interventions. To illustrate, outreach education involved costs relating to facilitator training, including their time and room hire (preparation costs); the time delivering the outreach visit (delivery costs); and the time of practice staff participation in outreach sessions (receipt costs). The fixed costs of the interventions are provided in *Table 11* (further details available on request). Our assumptions were generally conservative.

We assumed the fixed costs were equally distributed across all four ASPIRE implementation packages. However, to retain a conservative stance in our base-case analysis, we doubled the value of this per intervention cost when computing the fixed cost per practice ($n = 144$) of implementing each specific intervention. We undertook sensitivity analysis on the magnitude of the risky prescribing implementation package fixed cost. The base-case fixed cost was £2439 (£175,592.36/144 × 2) per practice. We tested an optimistic scenario where the per-practice fixed cost was £1219 (£175,592.36/144).

Variable implementation package costs included additional GP consultations to review and change patients' prescriptions, additional tests and the cost of additional medication (e.g. coprescribing gastroprotection). The variable cost element of ASPIRE was directly related to implementation package effectiveness in changing care processes; thus, no additional variable costs were incurred if GPs did not change practice on intervention receipt. Implementation package effects were measured using the trial arm differential in the proportion of eligible patients for each relevant indicator achieving recommended care or outcomes.

TABLE 11 Economic evaluation: fixed costs of ASPIRE implementation package (WP4b)

Fixed costs (£)	Educational outreach	A&F	Computerised tools
Intervention preparation	29,773.88	43,122.53	1459.32
Intervention delivery	39,269.49	7980.16	0.00
Intervention receipt	29,589.87	24,397.12	0.00
Subtotal	98,633.23	75,499.81	1459.32
Overall total ($n = 144$ practices)	175,592.36		

Risky prescribing

Our evaluation included all of the risky prescribing subindicators except for indicator nine, which covers the prescription of a diuretic in combination with an angiotensin-converting-enzyme inhibitor or angiotensin receptor blocker in those taking an oral NSAID. We could not find an acceptable evidence base on which to base parameters and it was relevant for only 3% of the current sample.

We based the risky prescribing model around a previous model of NSAID prescription used for the NICE osteoporosis guideline (CG177¹⁰⁴) (Figure 2). We recreated and adapted this model and populated it using results reported for three of the most commonly prescribed NSAIDs (diclofenac, naproxen and ibuprofen) and aspirin. Patients receiving NICE adherent prescriptions (e.g. a proton pump inhibitor with NSAIDs where indicated) are less likely to experience negative health events such as a gastric bleed. We assumed that patients experiencing an event would stay in the post-event health state for the remainder of the model time horizon. We assumed that patients in the post-event state had their NSAID stopped and were prescribed nothing or a simple analgesic such as paracetamol. The model has 3-month cycles and a lifetime duration.

The risk of adverse events per treatment were adjusted down to reflect lower-dose prescriptions in the UK. We estimated baseline adverse event risks by applying the following weights to each NSAID treatments based on data reported in the NICE guideline (see *Report Supplementary Material 1, Table 87*): diclofenac, 0.459; ibuprofen, 0.248; and naproxen, 0.077. *Report Supplementary Material 1, Table 88*, includes the additional risks associated with risky prescribing identified with targeted literature searches. *Report Supplementary Material 1, Table 89*, includes the effectiveness values used and *Report Supplementary Material 1, Table 90*, includes the model costs. We assumed variable intervention costs comprising two GP

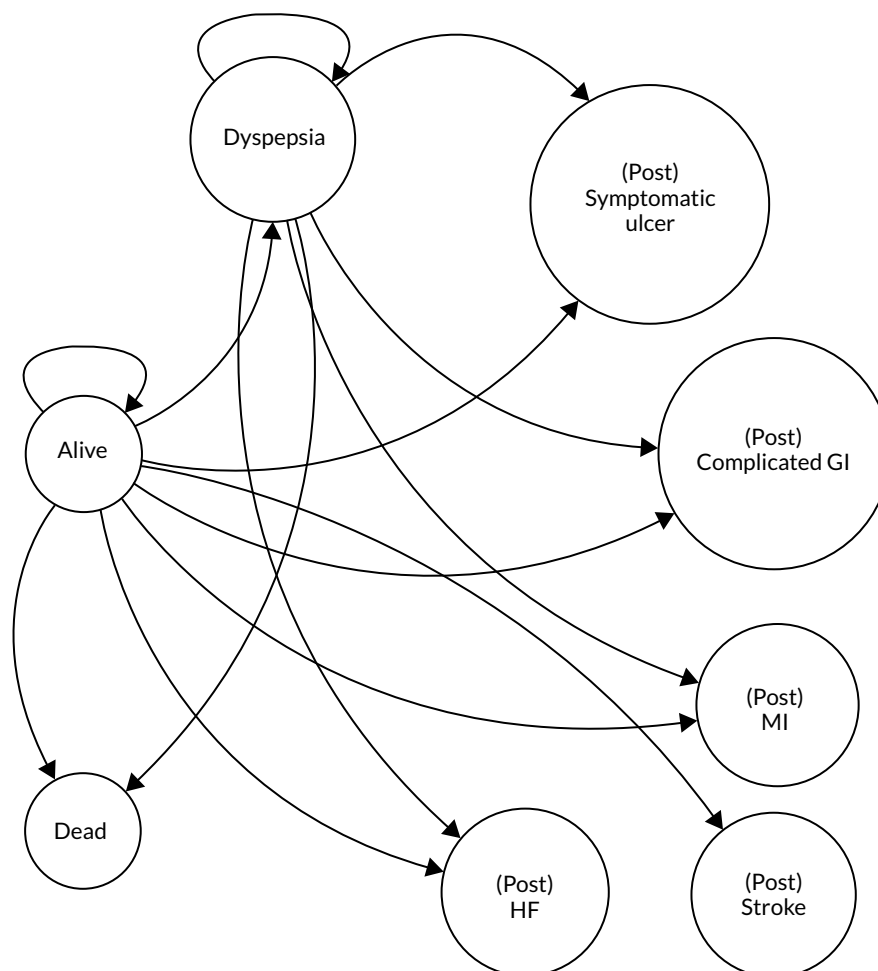


FIGURE 2 Economic evaluation: risky prescribing model structure (WP4b). GI, gastrointestinal; HF, heart failure.

visits (an initial consultation and one follow-up, £72 in total) and a proton pump inhibitor prescription where relevant (i.e. indicators 1–5 and 7). Given that changes in prescription may be made at other GP visits or by another means (e.g. telephone), our assumption of two GP visits is conservative.

As in the NICE guideline, the impact of adverse gastrointestinal and cardiovascular events on individuals' quality of life was modelled via utility multipliers (*Report Supplementary Material 1, Table 91*) applied to age-specific baseline utility scores for the general UK population.¹⁰⁵ The UK population baseline scores used were 0.78 [standard error (SE) 0.26] at age 65 years and 0.73 (SE 0.27) at age 75 years. With the exception of stroke, we assumed utility would return to pre-event levels after one cycle.

We created an additional model, based on that used to inform the NICE acute kidney injury (AKI) guideline (CG169),¹⁰⁶ to capture the risky prescribing indicator (9) covering the prescription of NSAIDs in patients with CKD (*Figure 3*). Transition probabilities are included in *Report Supplementary Material 1, Table 92*, with costs in *Table 93*. The variable intervention cost was assumed to be two additional GP consultations to review and change NSAID prescriptions (£72 per patient). *Report Supplementary Material 1, Table 94*, includes utility values, the same as used in CG169.¹⁰⁶ *Report Supplementary Material 1, Table 95*, includes the proportions used to weight the results.

Blood pressure control

The economic evaluation broadly followed the above methods for risky prescribing. We identified but could not obtain the model¹⁰⁷ informing the BP indicator (CG127);¹⁰⁸ therefore, we recreated and adapted it (*Figure 4* summarises the model structure). We based the modelling approach on that used for the NICE guidance for hypertension.¹⁰⁸ The model had a lifetime horizon and 3-monthly cycles with future costs and benefits discounted at 3.5% per annum. The cost and risks of a health event are often higher immediately post event and often decrease as time passes. To reflect this, all health states in the model have an 'acute' health state and a post-event health state. In all cases, if patients survive, they move automatically to the post-event health state after a period of 3 months (one cycle). The post-event health states are generally less costly and have reduced mortality risks.

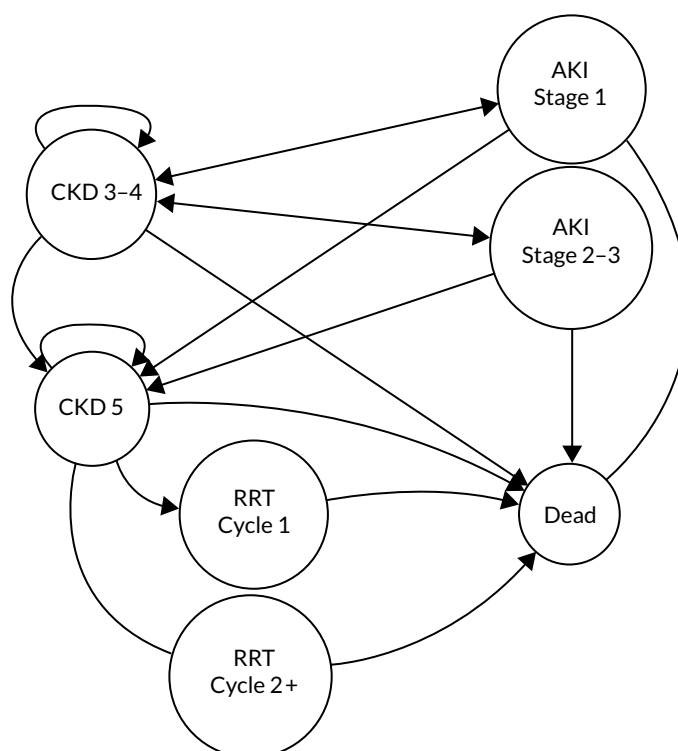


FIGURE 3 Economic evaluation: risky prescribing submodel – CKD (WP4b). RRT, renal replacement therapy. RRT are tunnel states only.

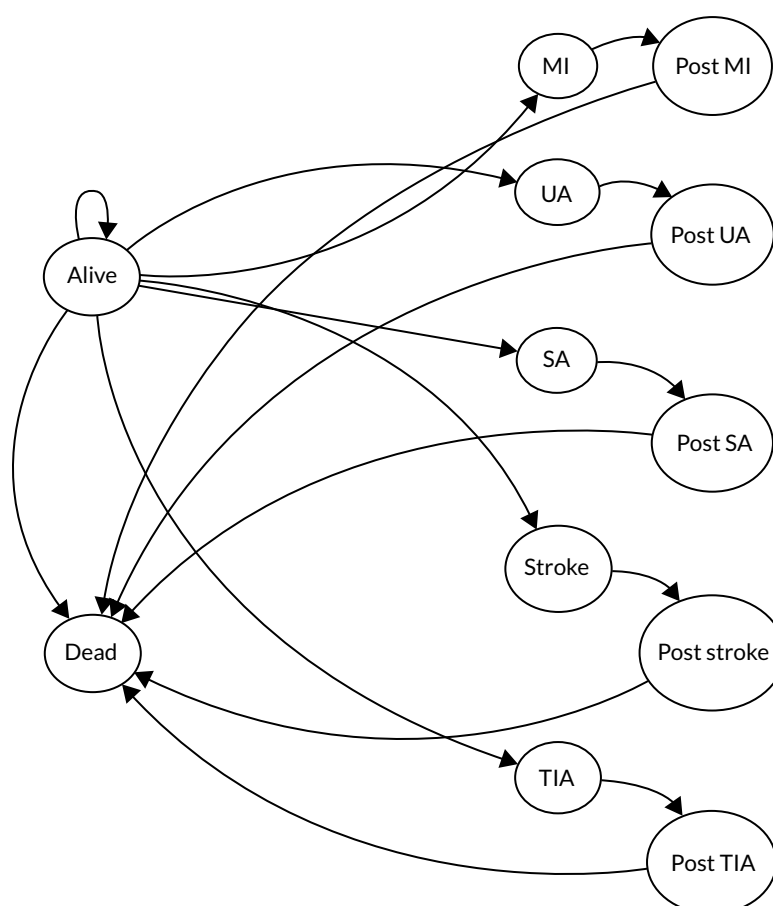


FIGURE 4 Economic evaluation: BP control model structure (WP4b). SA, stable angina; TIA, transient ischaemic attack; UA, unstable angina.

Report Supplementary Material 1, Table 96, includes the modelled effectiveness estimates and *Table 97* details transition probabilities. The risks for coronary heart disease and stroke are taken from the NICE CG127 model.¹⁰⁸ The original model based these risk parameters on outputs from the Framingham coronary heart disease and stroke equations by gender and age groups.¹⁰⁹ We employed these values and simulated associated CIs using a risk calculator spreadsheet developed by the University of Edinburgh.¹¹⁰ Although the Framingham equations have been generally superseded in the UK by QRISK,¹¹¹ we mirrored the original modelling approach. As the risk equations do not provide more granular detail about type of coronary heart disease and stroke, we followed NICE CG127¹⁰⁸ modelling and used percentage distributions across coronary heart disease (MI, stable angina, unstable angina) and stroke [stroke or transient ischaemic attack (TIA)] event types.¹¹² Although some health-event risks described may be correlated (e.g. stroke and MI), the NICE approach assumes the events are independent; we followed this assumption. Patients meeting treatment targets for hypertension have a lower risk of coronary heart disease and stroke. Intervention effectiveness was determined by mean, adjusted, systolic BP outcomes at trial end. The relative risk reduction for male and females following treatment receipt were taken from a meta-analysis.¹¹³ To cover all indicator populations, we identified additional risk parameter values using targeted searches and risk calculators. However, we used the same background mortality (UK life tables) and BP control risk reduction values for all indicators.

We based the average cost of hypertension treatment on antihypertensive drug therapy and an annual check with the GP (see *Report Supplementary Material 1, Table 98*). As we had no data on changes in process of care, we explored alternative costs associated with differing levels of implementation. We took costs of cardiovascular events from the NICE guideline,¹⁰⁸ national cost sources and published studies. As in the original evaluation, we assumed that these costs were fixed and thus not random in

the probabilistic sensitivity analysis. We based the quality-of-life (utility) values (see *Report Supplementary Material 1, Table 99*) for the cardiovascular event health states on the NICE statins assessment. We assumed no utility decrement for receipt of hypertension treatment or experience of a TIA. *Report Supplementary Material 1, Table 100*, includes the proportions used to weight the results.

Diabetes

We used the United Kingdom Prospective Diabetes Study (UKPDS) outcomes model,⁹⁸ which had informed the NICE diabetes guideline (NG28)¹¹⁴ and is a comprehensive outcomes model based on a large, UK longitudinal data set. The model produces estimates of life expectancy, quality-adjusted life expectancy and cost of therapies and complications. Confidence intervals were created using the bootstrapping technique. This individual-level model requires patient-level data on a range of demographics (e.g. age and ethnicity), risk factors (e.g. body mass index and smoking status) and existing health conditions (e.g. amputations and blindness). As the trials did not provide patient-level data, a cohort of individuals were simulated using the mean and CIs from the trial analyses. For simplicity, and in the absence of other data permitting an alternative approach, we assumed these to be normally distributed. We randomly generated data on other model input parameters (e.g. gender, smoking status and presence of AF) such that they satisfied prevalence estimates in this group (e.g. such that 3.9% of patients had existing AF; see *Report Supplementary Material 1, Table 101*). As correlations between these input parameters were unknown and uncontrolled for, the simulations are likely to lead to more rather than less uncertainty.

A sample of 1000 patients was simulated for each arm, the only difference in the sample being trial effectiveness data (systolic BP, low-density lipoprotein cholesterol and HbA_{1c}), simulated based on trial results (and their CIs). We captured first order uncertainty by simulating outcomes for each patient multiple times in loops and second order (sampling) uncertainty was captured by bootstrap resampling of the 1000 patients. One hundred loops and 1000 bootstraps were conducted. As a sensitivity analysis, we also ran the model for 1000 loops and no bootstraps, including only the mean outcome values for the two trial arms. In the absence of other data to inform dynamic parameter values, we carried forward baseline model input values into each model year. The fixed cost of the intervention at a patient level was £0.28 [(£175,592.36/178 × 2)/7130]. As the outcomes from the trial entered in the model were continuous health outcomes, it was not possible to attach to them costs associated with a change in practice (process outcome); however, we explored this using sensitivity analyses in which additional resource was expended in the trial arm.

Results

Risky prescribing

For an average UK practice of 7130 patients, the implementation package targeting all risky prescribing indicators (except indicator 9) is more expensive but more effective than standard care (*Table 12*). The risky prescribing implementation package ICER is below the threshold (£20,000) used by NICE. Although the values are per practice, the ICER of £1359 is also that per patient and generally represents comparative value.

TABLE 12 Economic evaluation: mean outcomes for the composite risky prescribing indicator (WP4b)

Mean incremental QALY	0.902
Mean incremental cost	£1225
ICER	£1359
Incremental net monetary benefit at £20,000 per QALY	£16,810

Figures 5 and 6 show the probability sensitivity analysis (PSA) results in a scatter plot and a CEAC, respectively. The spread of the simulated ICERs on the plane suggest greater uncertainty in terms of incremental QALYs than in incremental costs. A greater proportion of ICERs fall below than above the cost-effectiveness threshold line. The probability that the implementation package ICER is below £20,000 is relatively high, chiefly because of low intervention costs; hence, even at a low benefit, it is likely to represent value for money. At a QALY willingness-to-pay (WTP) threshold of £20,000, the risky prescribing intervention has a > 79% chance of being cost-effective.

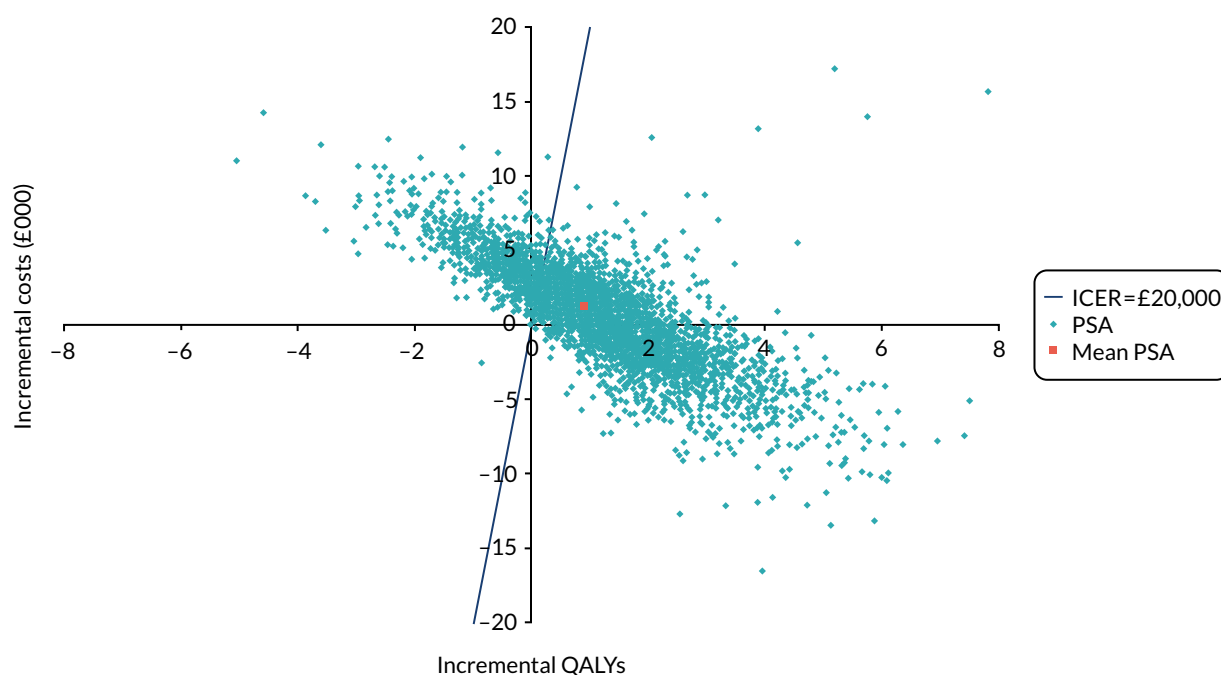


FIGURE 5 Economic evaluation: composite indicator of risky prescribing – scatterplot of simulated incremental cost and QALY (WP4b).

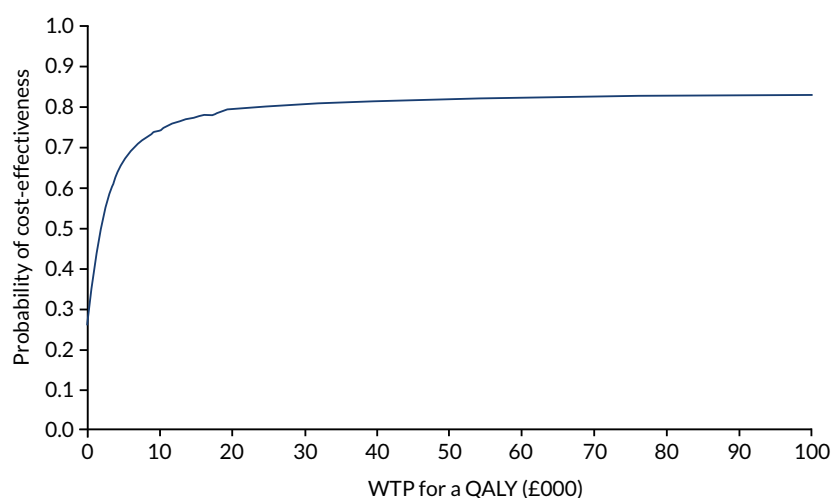


FIGURE 6 Economic evaluation: composite indicator of risky prescribing – CEAC.

Results vary greatly by indicator. *Figures 7–10* contrast cost-effectiveness results for indicator 4, where the trial-arm difference in prescribing is statistically significantly better in the implementation package arm; and for indicator 2, where the prescribing was better in the control arm. For indicator 4, a far greater proportion of simulated ICERs are below the cost-effectiveness threshold. *Figure 11* is a CEAC run when the fixed intervention costs are adjusted to alternative optimistic (cheaper) and conservative (more expensive) scenarios. Even using a conservative costing scenario, the risky prescribing implementation package has a > 70% chance of being cost-effective.

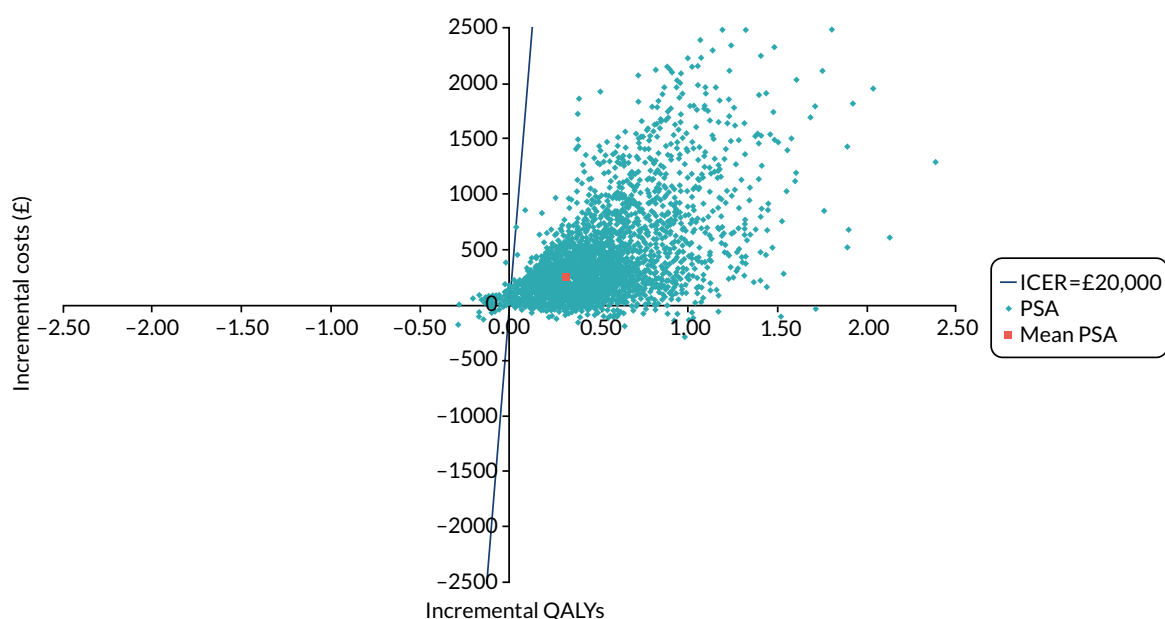


FIGURE 7 Economic evaluation: risky prescribing indicator 4 (prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of gastroprotection) – scatterplot of simulated incremental cost and QALY (WP4b).

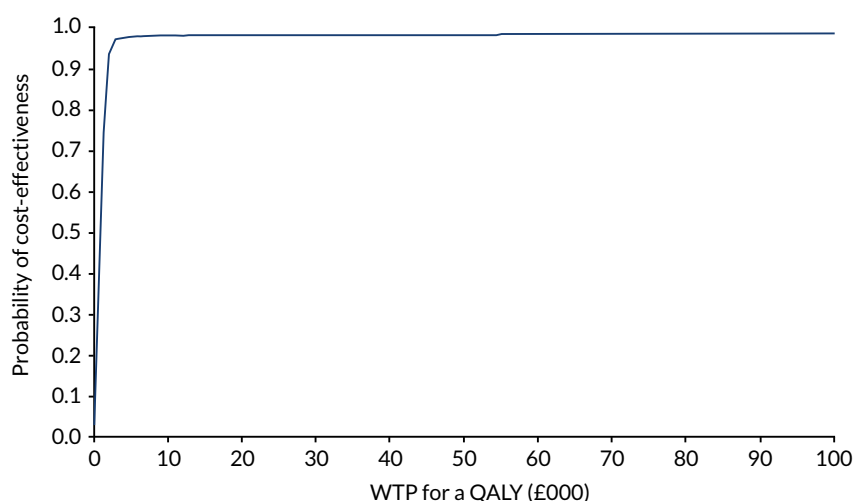


FIGURE 8 Economic evaluation: risky prescribing indicator 4 (prescribing of aspirin and clopidogrel in patients aged ≥ 65 years without coprescription of gastroprotection) – CEAC (WP4b).

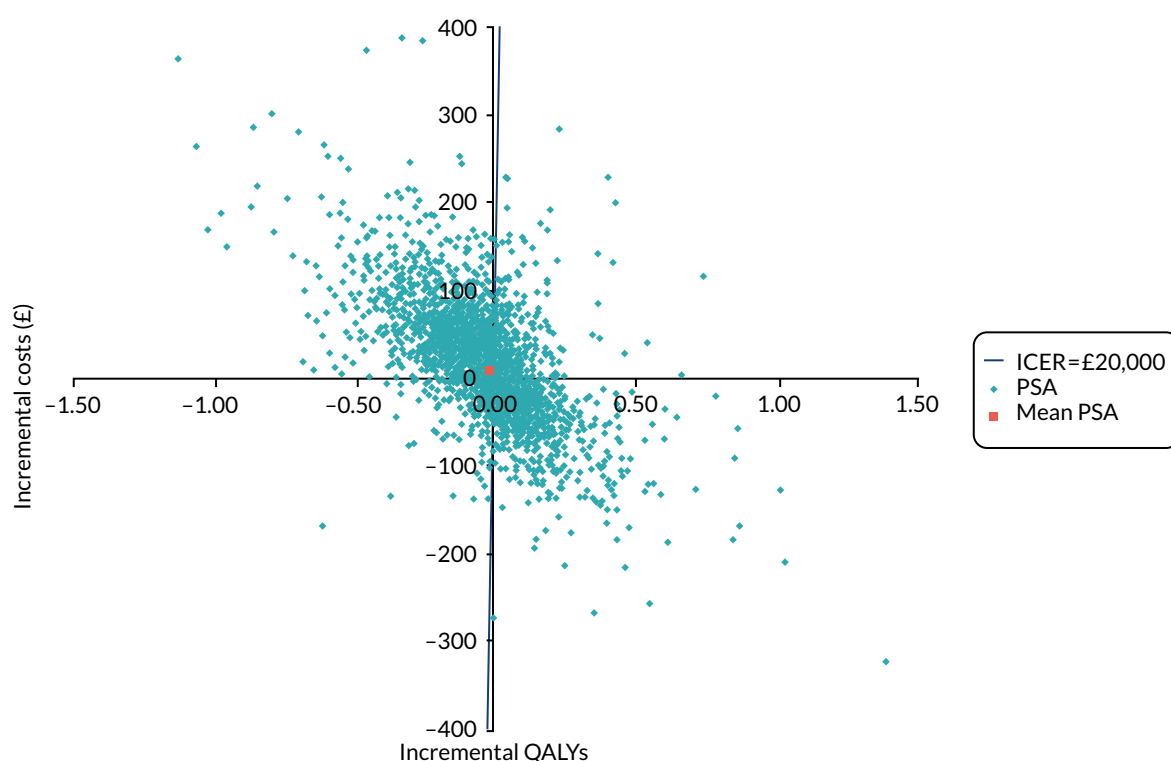


FIGURE 9 Economic evaluation: risky prescribing indicator 2 (prescribing of traditional oral NSAID in patients aged ≥ 75 years without coprescription of gastroprotection) – scatterplot of simulated incremental cost and QALY (WP4b).

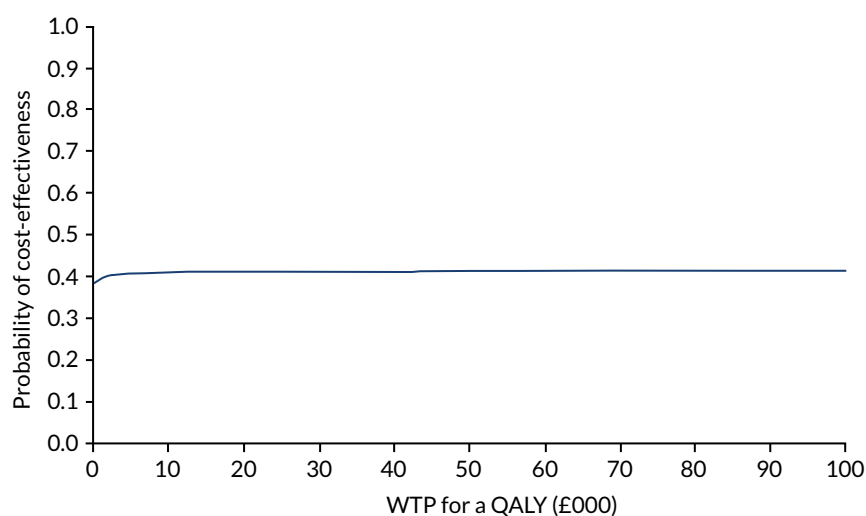


FIGURE 10 Economic evaluation: risky prescribing indicator 2 (prescribing of traditional oral NSAID in patients aged ≥ 75 years without coprescription of gastroprotection) – CEAC (WP4b).

Blood pressure control

The BP implementation package at practice level was more expensive (incremental cost = £42,192) but more effective (incremental QALYs = 13.00) than standard care (*Table 13*). The incremental values at patient level were very small (£5.92 for costs and 0.0018 for QALYs). The estimated ICER per practice and patient was £3246, which is indicative of cost-effectiveness; however, there is considerable uncertainty in the results.

Figures 12 and 13 show the PSA scatter plot and CEAC, respectively. A slightly greater proportion of ICERs fall below than above the cost-effectiveness (£20,000 per QALY) threshold line. This is reflected in the finding that the probability of the intervention being cost-effective is only just above 50%, at 51.9%.

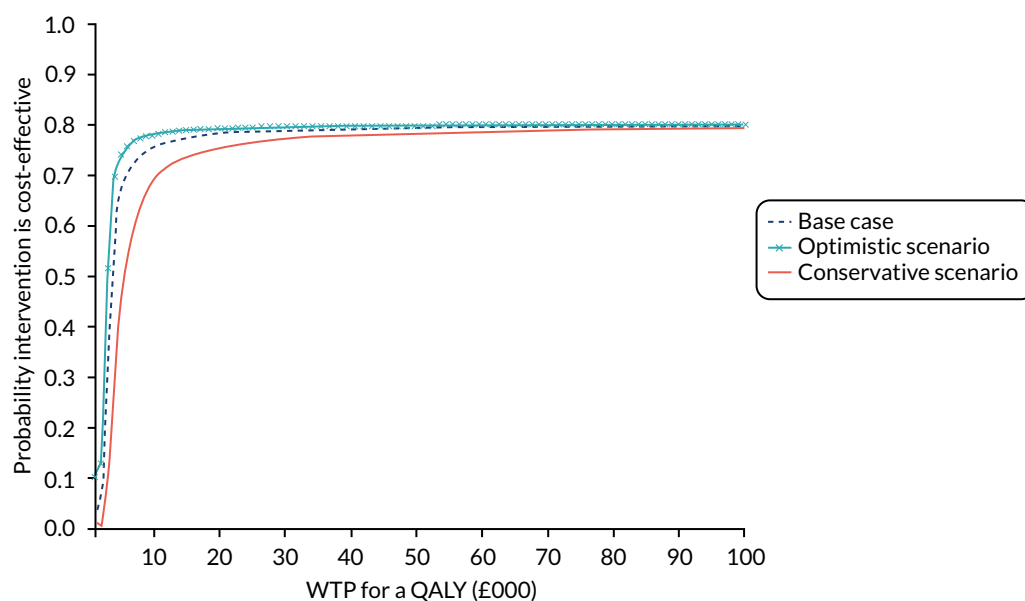


FIGURE 11 Economic evaluation: composite indicator of risky prescribing – CEAC for three sets of fixed intervention costs (WP4b).

TABLE 13 Economic evaluation: mean PSA outcomes at the practice level for the composite BP control indicator (WP4b)

Mean incremental QALY	13.00
Mean incremental cost	£42,192
ICER	£3246
Incremental net monetary benefit at £20,000 per QALY	£217,730

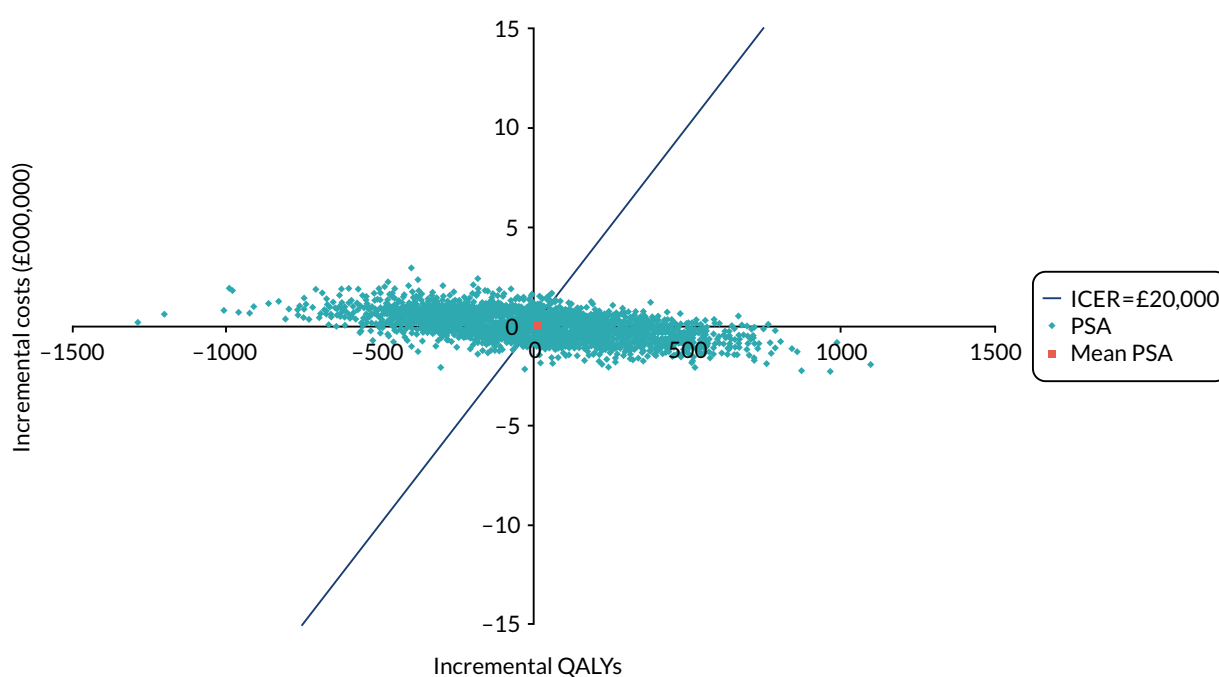


FIGURE 12 Economic evaluation: composite indicator of BP control – scatterplot of simulated incremental cost and QALY (WP4b).

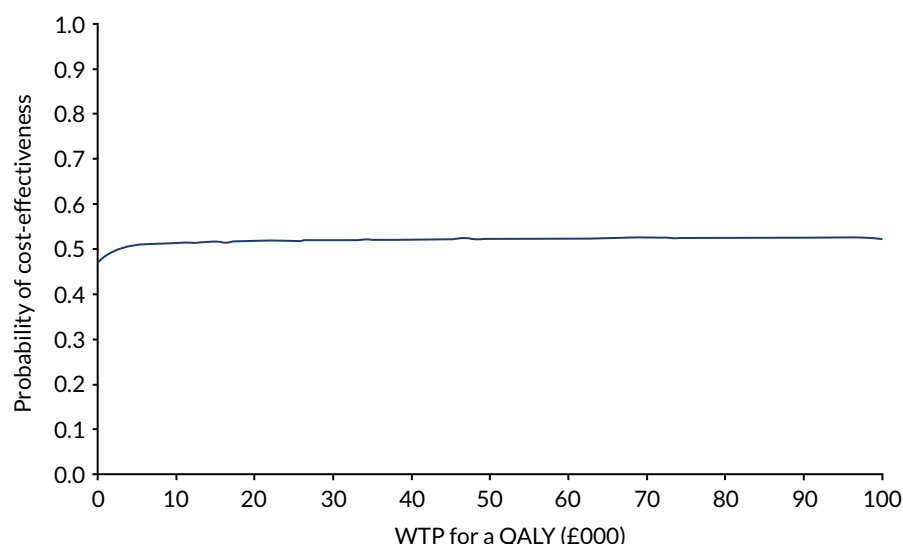


FIGURE 13 Economic evaluation: composite indicator of BP control – CEAC (WP4b).

Results by subindicator are largely similar with the intervention yielding positive net benefit, either dominating standard care or yielding an ICER of < £6000 per QALY gained. Two exceptions were the indicators relating to patients with a history of stroke or TIA and relating to patients with diabetes for whom standard practice (no intervention) was more cost-effective than the BP implementation package. This was as a result of greater mean reductions in BP in the control arm. In a third indicator (peripheral arterial disease) the outcomes were practically the same across arms. In the best performing indicator (9.5: BP control in patients with CKD), the PSA indicated a 63% chance of being cost-effective (Figures 14 and 15).

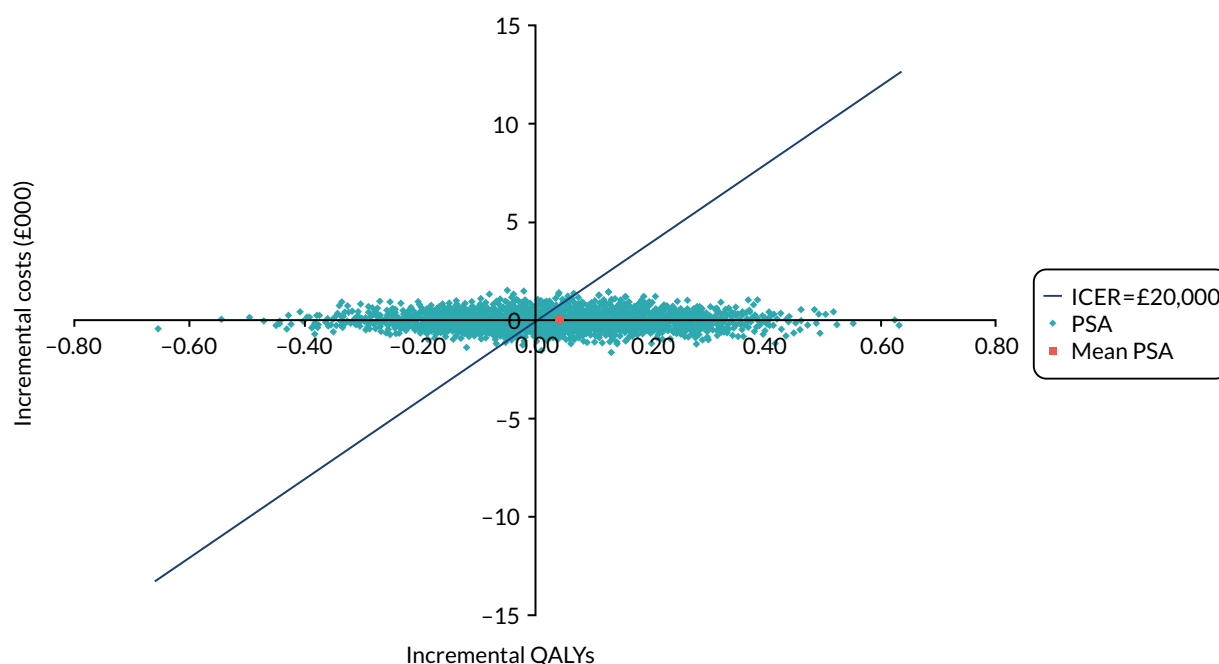


FIGURE 14 Economic evaluation: BP control indicator 5 (achievement of BP < 140/90 mmHg in patients aged < 80 years with peripheral arterial disease) – scatterplot of simulated incremental cost and QALY (WP4b).

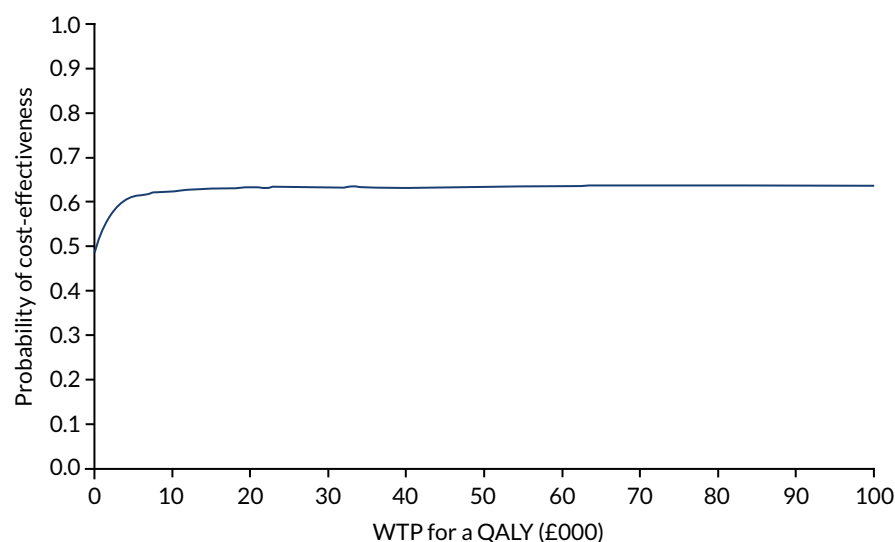


FIGURE 15 Economic evaluation: BP control indicator 5 (achievement of BP < 140/90 mmHg in patients aged < 80 years with peripheral arterial disease) – CEAC (WP4b).

Diabetes control

Table 14 includes the mean of the individual-level summary results from the UKPDS model. The differential between trial arms in costs and benefits was negligible. Over the model's lifetime horizon, 12.26 and 12.24 life-years were generated by the implementation package and control arms, respectively. This equated to 9.78 and 9.76 QALYs, respectively; thus, the package appeared more effective than standard practice (incremental QALYs = 0.02). The implementation package was more expensive overall (incremental cost £12.79); these incremental values yielded an ICER of £573.37. At WTP thresholds of £20,000 and £30,000 per QALY, the implementation package has only a 34% and 35% chance, respectively, of being cost-effective. The spread of ICERs from the bootstrapping is shown in Figure 16; however, the Monte Carlo Error (MCE) for the initial model (100 loops and 1000 bootstraps) was of the same magnitude as the incremental values indicating highly uncertain results. A second model run where the sample outcomes were simulated over 1000 loops (0 bootstraps) indicated incremental QALYs of -0.0007 and costs of £6.37 over a lifetime (control dominates). A further run including only mean values for the arms over 10,000 loops (0 bootstraps) generated incremental QALYs of -0.0085 and costs of £28.04.

TABLE 14 Economic evaluation: mean individual-level summary results from the UKPDS model (WP4b)⁹⁸

	Life expectancy (years)		QALYs		Complication costs (£)		Total cost (£)
	Mean	95% CIs	Mean	95% CIs	Mean	95% CIs	Mean
With 1st order uncertainty							
ASPIRE	12.26	12.10 to 12.64	9.78	9.65 to 10.07	24,420.87	24,043.73 to 25,124.88	24,421.15
Usual care	12.24	12.11 to 12.65	9.76	9.65 to 10.10	24,408.36	24,027.83 to 25,103.85	24,408.36
Incremental		0.02		0.02		12.51	12.79
ICER							573.37

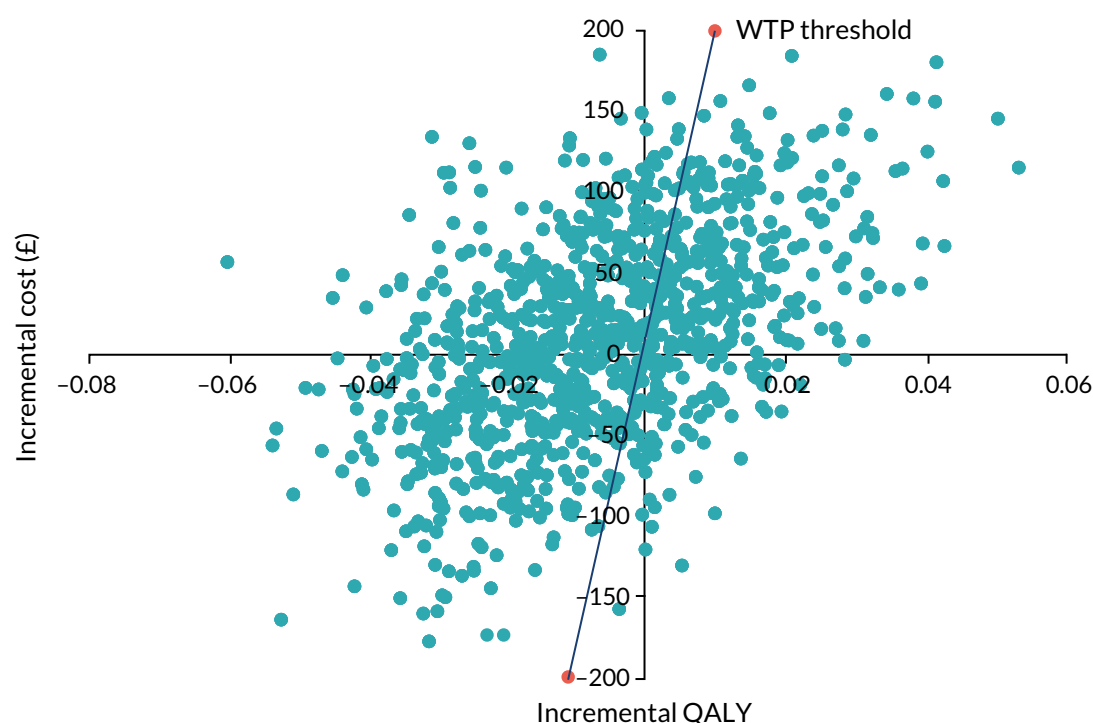


FIGURE 16 Economic evaluation: cost-effectiveness plane for diabetes control implementation package (WP4b).

Given these outcomes, their sensitivity to model processes and the magnitude of differences observed, we cannot be confident in the results and we cannot recommend the diabetes implementation package. As the intervention costs were negligible, and there was no reasonable argument for adjusting the effectiveness parameters input into the model, sensitivity analyses were not conducted.

Discussion

Although the need to consider the economic consequences of implementation is acknowledged,¹¹⁵⁻¹¹⁷ full economic evaluations of implementation interventions are relatively rare. We undertook economic evaluations addressing three of the four ASPIRE implementation packages. Where possible, we followed the reference case set out by NICE⁹⁹ and applied the same modelling approaches used to inform the development of the original NICE guidance. We conducted detailed costing of the preparation, delivery and receipt of different components of the implementation packages. At a practice level, the cost for all four packages was low (£1219.39) and at a patient level negligible (£0.17). Details on the costs may facilitate wider commissioning and uptake of such interventions.

The risky prescribing implementation package was, on average, more expensive and more effective than usual care. This yielded an ICER of £1359, well below the £20,000-per-QALY threshold. This was supported by the PSA in which > 75% of simulations yielded ICERs below this threshold; however, the intervention benefit was not evenly spread across subindicators and some (e.g. subindicator 4) drove the results.

The BP implementation package was also more expensive and slightly more effective than usual care with the evaluation yielding an ICER of £3246. Although this is indicative of cost-effectiveness, the uncertainty around this result was considerable and only 50% of the ICERs fell below this threshold in the PSA.

There was mixed trial evidence for the effectiveness of the diabetes implementation package where it was observed to be either minimally more, or less, effective than usual care, depending on the clinical end point. This led to minimal differences in costs and benefits and the conclusion that the package was unlikely to be cost-effective.

Given that there were so many elements to the models, it was not possible to systematically vary parameters deterministically in the analysis; hence, apart from some sensitivity analyses based on intervention costs, we relied on the probabilistic analyses to inform on the uncertainty in the results. In general, the results of the evaluations were insensitive to changes in costs given that these were negligible. The results indicate that although effects may be small and statistically non-significant in some cases, the relatively low cost of the implementation packages (especially around risky prescribing and possibly around BP control) suggests that they are worthy pursuits and may represent value for money. Value for money may even be underestimated if any intervention effects can be sustained beyond trial follow-up periods.

A limitation of the evaluations is their reliance on decision modelling rather than an individual-level data analysis; however, it is unclear whether or not access to individual data would have changed results and would have been worth the added complexity.

Conclusion

Economic evaluations alongside implementation studies are relatively uncommon but are important to inform priority setting and resource allocation. We have demonstrated that there is a good probability of the risky prescribing implementation package representing value for money but we would be more cautious in recommending the BP control package.

Work package 5: a multimethod process evaluation of an adaptable implementation strategy

Background

The randomised evaluation of the adaptable implementation package found a cost-effective reduction in risky prescribing but no effects on diabetes control, BP control or anticoagulation for AF. Our process evaluation aimed to help explain these findings. We created chronological accounts of how the implementation package was delivered, received and acted on. We applied the TDF⁵⁴ to conceptualise planned intervention content and implementation (i.e. the theoretical domains targeted by the intervention and likely to influence processes of care and outcomes). Normalisation process theory (NPT) offers a sociological approach to understand the dynamics of implementing, embedding and integrating a new technology or complex intervention.^{118,119} It is concerned with explaining what people do rather than their attitudes or beliefs. It explains actions according to a set of four constructs; coherence, cognitive participation, collective action and reflexive monitoring. We applied NPT to help identify implementation processes in practice teams over time. We developed process models to explore how actual implementation compared with planned implementation.

Methods

Participants

Eight general practices not participating in the trials took part in the in-depth, mainly qualitative process evaluation. An independent statistician randomly assigned practices to one of four adapted implementation packages, which were the same as those used in both trials. Allocation was balanced to reflect CCG membership and practice list size. We gathered further fidelity data from all practices assigned to the implementation package in the trial.

Field work (Table 15) with the process evaluation practices was conducted by Cheryl Hunter, a social scientist researcher. Her role was to observe and collect data but not to interfere with implementation package delivery, as far as possible.

TABLE 15 Summary of fidelity data collection methods (WP5)

Intervention components	Data collection methods	Data collected from
A&F, including reports, computerised searches, significant event audit forms	Post-trial survey (receipt and use of reports and searches)	All practices
	Record of educational outreach visit	All practices
	Record of joining organisational group (to access searches)	All practices
	Observations and interviews	Process evaluation practices
Initial educational outreach visit, including outreach support	Record of visits from facilitators	All practices
	Record of reasons for declining visit	All practices
	Observation of visits/support	Process evaluation practices
	Interviews with practice staff	Process evaluation practices
	Interviews with facilitators	All practices
Prompts, including computerised protocols (for risky prescribing arm only) and physical reminders (laminates, pens and sticky notes)	Record of visits from facilitators	All practices
	Observations and interviews with practice staff	Process evaluation practices
	Record of joining organisational group (to access protocols)	Risky prescribing practices

Reproduced from Glidewell *et al.*⁷¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Interviews

Cheryl Hunter carried out theory-based semistructured interviews (informed by the TDF and NPT constructs) with key members of practice staff at two time points. These included clinical leads, practice managers and other staff involved in the organisation or delivery of care for the specific QI. The first interviews followed delivery of the initial educational outreach session and explored the role and responsibilities of the staff member, barriers to and facilitators of achieving the QI, responses to the intervention components and subsequent actions by the practice (35 participants; 3–6 per practice). The second interviews took place towards the end of the intervention year and revisited questions from the first interview as well as exploring changes within the practice and perceived intervention usefulness and fit (27 participants; 2–4 per practice). Interviews were audio-recorded, transcribed and anonymised.

Cheryl Hunter conducted a final group interview at all eight practices 1–3 months after delivery of the final feedback report. She asked practices to reflect on the entire year, provide feedback on intervention components and explore any further plans to conduct work around their assigned QI. Group interviews were written up in detailed anonymised field notes.

Cheryl Hunter conducted interviews with nine facilitators who delivered the educational outreach visits and outreach support. These interviews explored how outreach sessions and support were delivered in practice, whether or not the facilitators felt that these were delivered as intended and what seemed to work (or not) when delivering these components in practice.

Observations

Approximately 10 hours of observation was conducted over the year at each practice. Observations were informed by TDF and NPT constructs (e.g. attending to staff, individually and collectively, conceptualise the intervention and what barriers or enablers might contribute to implementation). Cheryl Hunter attended ASPIRE-related meetings, such as educational outreach sessions and on-site outreach support. Cheryl Hunter also observed routine practice in reception or administrative areas and attended a variety of practice meetings (including peer review, nursing team, clinical and business meetings). No clinical interactions, such as consultations, were observed. Anonymised field notes structured around the TDF and NPT constructs were created of all meetings and conversations at practices.

Documentary data

We collected related documents including protocols for the assigned QI, letter templates for patients, patient information leaflets and minutes from relevant clinical or practice meetings. E-mail correspondence and notes from telephone conversations between Cheryl Hunter and practice staff were anonymised. Practice managers were also given box files and invited to use these to store any documents relevant to the QI.

Intervention administrative data

Feedback reports were sent quarterly by post and e-mail, and copies were taken by facilitators to outreach visits. E-mail delivery receipts were recorded. For the eight practices that participated in the in-depth process evaluation, delivery, receipt and use of feedback reports, searches and significant event audit forms were tracked through observations, interviews and documentary records. Practices were invited to join a SystmOne organisational group and were reminded about it in feedback reports, by e-mail and at outreach visits. We collected data on whether or not practices joined the group (which gave automatic access to the computerised searches), but it was not possible to digitally track use of the searches. All practices were invited to participate in outreach visits by a programme administrator, who recorded data on reasons for declining outreach visits. The facilitators delivering educational outreach and support also recorded who attended outreach sessions, how many sessions were delivered, what additional pharmacist support was taken up and delivered to practices, and whether or not feedback reports were received and used. For risky prescribing only, practices were given access to computerised prompts through an organisational group (in the same manner as the computerised searches). Further insight into the delivery and receipt of the risky prescribing computerised protocols came from interviews and observations with two process evaluation practices allocated to this intervention. Laminates, pens and sticky notes summarising key clinical messages were delivered by post throughout the year (and sometimes via facilitators at outreach visits). Interviews and observations at the eight practices gave further insight into intervention receipt and use.

Post-trial survey

We aimed to estimate intervention fidelity post trial in both trial and process evaluation practices. We developed a brief e-mail survey based on NPT to explore intervention receipt, enactment and ongoing workability. The survey asked whether or not the practice received the reports, the reports were relevant to the practices, the reports were shared with colleagues in the practice, the reports were discussed within the practice, the reports were used to change how people in the practice worked and they had used the SystmOne searches. Response options were limited to 'yes' or 'no', but practices could also provide further free-text feedback if desired.

Analysis

Interview and observational data were entered into a QSR NVivo 10 project and coded using a framework approach.^{54,120,121} The analytic framework included TDF constructs to compare planned with actual implementation and NPT was adopted to understand individual- and team-implementation processes, in particular how individuals and groups conceive of, engage with, and enact and reconfigure work in response to the implementation package. Data were grouped by practice and chronologically

ordered. Review of coded data, chronological accounts and process models was carried out in sequential team meetings. The team comprised researchers with experience in psychology, sociology, implementation science and primary care.

Cheryl Hunter developed a logic model (see *Figure 18*) and process models (see *Figure 19*) for each intervention delivery mechanism to describe the planned process of implementation as conceived by the intervention developers. Cheryl Hunter also drew up 'disrupted' process models, highlighting hypothesised points where difficulties could arise to disrupt implementation as planned. Process and 'disrupted' process models were used to sensitise the researcher and analysis team during data collection and analytic discussions. Process models depicting implementation and chronological narrative accounts were developed for each process evaluation practice documenting intervention delivery, receipt and use. The post-trial fidelity survey coded fidelity as high, medium and low. Fidelity was considered high if practices received feedback reports, took up outreach and accessed patient-identifiable searches; medium if they received feedback and outreach or searches; and low if they received only feedback reports. Survey and outreach facilitator data from the trials were analysed descriptively and compared with the more in-depth practice case narratives, to help explain the trial findings and augment insights into implementation fidelity.

Results

We initially outline the pre-intervention context before describing the delivery, receipt and enactment of the main implementation package components as well as how they appeared to work or not. We then summarise how or whether or not the general practices appeared to change as a result of the package. Our descriptions relate to the in-depth process evaluation practices unless stated otherwise. *Table 26* in *Appendix 1* summarises findings related to delivery and receipt of the implementation package, perceptions of package components and actions in response to the package.

Pre-intervention context

Pre-intervention achievement was broadly comparable between trial and process evaluation practices across trials and indicators, with any variations reflecting the smaller sample of process evaluation practices (see *Report Supplementary Material 1, Table 83*).

The practices had varying systems and procedures in place for targeted indicators. The most advanced were generally for diabetes, for which there was a nominated lead GP and care was mainly shared between nurses and GPs. Patients were recalled for reviews at intervals according to their progress in attaining treatment goals. Care was organised similarly, if less intensively, for BP management. BP monitoring was variously carried out by patients using home monitors, machines in waiting rooms or in consultations with health-care assistants, nurses or GPs, with the latter mainly having responsibility for prescribing. Practices had delegated GP leads for cardiovascular disease or AF. Neither practice assigned to the AF package involved their nursing staff in managing this condition. Both practices were aware of anticoagulation as a priority in their CCGs and of recent changes to clinical guidelines, and had started or had plans to start systematically reviewing patients who were potentially eligible for treatment. For risky prescribing, both practices assigned to this package had nominated GP prescribing leads. In one practice, a prescribing clerk regularly reviewed repeat prescriptions and did work around reviewing costs of prescribing for medicines management. The lead GP also met with the practice manager routinely to review any audits or work required for the CCG. Risky prescribing was the only targeted indicator without specific QOF-related systems of care and goals.

Intervention delivery and receipt

All trial and process evaluation practices received feedback reports as intended (except where there had been closures or mergers). Outreach visits were delivered to 67 (46.53%) practices, with eight (5.56%) receiving two visits. Uptake was similar between trial arms. Visit timings are presented in *Figure 17*. Detailed summaries are in *Appendix 1 (Tables 26–29)*.

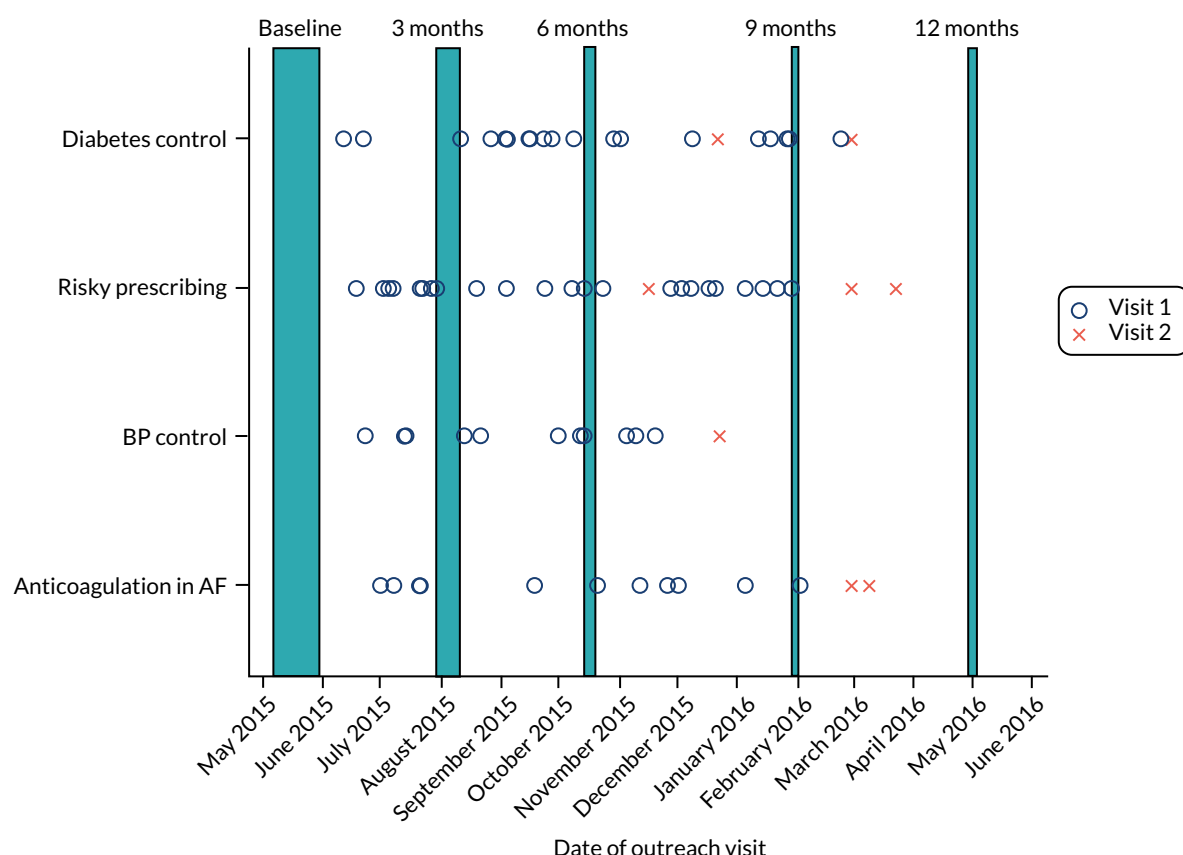


FIGURE 17 Cluster-randomised evaluation: distribution of delivery of outreach visits across the intervention period. Bars indicate delivery of feedback reports.

Outreach visits averaged 36 minutes (see *Report Supplementary Material 1, Table 77*). Risky prescribing and anticoagulation outreach meetings were all attended by someone with leadership responsibilities compared with BP control (81.82%) or diabetes control (85.00%). On average fewer practice staff attended risky prescribing meetings: three (range 1–10) compared with four (1–13) for anticoagulation, 4.5 (1–15) for diabetes control and 5.5 (2–14) for BP control (see *Report Supplementary Material 1, Table 78*). Key clinical topic leads were present at most visits (see *Report Supplementary Material 1, Tables 79 and 80*). Most practices (94.03%) that received outreach visits produced action plans (see *Report Supplementary Material 1, Tables 81 and 82*). During the outreach session, practices reported participating in a wide range of other quality improvement initiatives, particularly for diabetes (see *Report Supplementary Material 1, Tables 83 and 84*). Outreach visits were declined by 77 practices; 52 (67.53%) provided no reason and 22 declined for reasons related to workload (11 practices; 14.29%) or lack of interest (11 practices; 14.29%). One hundred and twenty-six (87.50%) practices joined the organisational groups (see *Report Supplementary Material 1, Table 82*; for baseline characteristics, see *Tables 85–86*), allowing them to access searches and, for risky prescribing, computerised prompts. Overall, fidelity receipt was high for 47 (32.64%) practices, medium for 82 (56.94%) and low for 15 (10.42%) (see *Appendix 1, Table 28*).

Audit and feedback

Process evaluation practices consistently viewed their QI as important and the data source as credible; however, practices typically interpreted the reports as relevant only to staff directly involved in clinical care for that indicator. This was demonstrated in the pattern of report dissemination and in how and whether or not reports were discussed as a group. Staff largely continued existing ways of working.

The feedback reports seemed motivational and to raise awareness of key clinical messages, but did not consistently encourage individual or organisational behavioural change. Required actions were generally interpreted as relevant only to one or two people, or seen as already aligned with existing

tasks and organisational structures. Actions were typically assigned to one or two GPs with little existing 'slack' and progress depended on whether or not and how they prioritised and managed required work; subsequently, progress was often slow and sometimes non-existent.

Some staff responded negatively to continuing feedback; this occurred in diabetes control and BP control practices in which the numbers to review were high (up to 400 patients) and considerable resources were already directed towards QOF work. Feedback in this case created dissonance, as practice staff felt overwhelmed yet saw no progress. They then queried the data or the outcomes measured as a way to resolve this dissonance and protect their professional sense of self.

One practice seemed to take a whole-practice approach to involvement, facilitated by their small size (< 10 staff members). Engaging their administrative staff kept the work going even when the GPs felt too busy, and the different staff members reminded and supported each other. This collective action, with clear roles for each staff member, may explain their steady improvement throughout the study.

The searches generated lists of patients requiring action by staff. However, the searches were not particularly visible to staff, suggesting a mismatch between the administrative and managerial staff who received them and the clinicians responsible for reviewing the generated patient lists.

Educational outreach and support

Practices tended to perceive educational outreach as the start of the intervention; they received the initial reports and then awaited the proffered help. Facilitators perceived these visits to be more useful when some practice staff had reviewed and considered the reports first; our observations suggested that only a small proportion of practices did this.

Who attended outreach visits was left to the practice's discretion. As noted earlier, practices generally had existing team arrangements for diabetes, involved everyone to some extent in BP control, and mostly had GP prescribers, with a named GP lead for prescribing. Practices also tended to fit the outreach visit into a routine meeting (as we had intended). This meant that staff awareness of the study and engagement with the educational outreach visit varied. It might have resulted in more tailored action plans if the diabetes and BP arms were given more guidance as to who to invite and include in the visits. Only practices in these two intervention arms provided any negative feedback on the outreach visits.

Outreach support seemed to create an expectation that someone else (the facilitator) would do the work for the practice. Some practices turned the outreach support down as they felt it better to do the work themselves; others accepted the offer because they lacked capacity to do the work themselves. The size of the workload was an influence, with diabetes and BP control workloads being viewed as overwhelming. Practices might have been more motivated if they had to focus on smaller target populations, or if plans had more clearly outlined individual roles and expectations. The clear link between action and outcome in the anticoagulation and risky prescribing practices facilitated engagement with action-planning and motivated staff to engage; the link between action and outcome was much less clear in the diabetes and BP control practices.

Delays between organising and delivering the outreach visits (see *Figure 17*) and support are likely to have differentially affected progress by QI. Practices assigned to diabetes and BP control packages typically required a number of staff to be involved in seeing patients (for tests, prescriptions and lifestyle advice), and would then require further review to establish any effect of management changes on outcomes. Arranging educational outreach sessions within the first 6 months of the trial was not feasible for many practices. If the ostensible 'start' of the intervention work was after the outreach visit, patients might not have been identified for action until near the end of the trial intervention period.

Computerised prompts and laminated reminders

Observations in the process evaluation practices suggested that the laminated reminders had low visibility and were unlikely to have contributed to any intervention impact. The only observed indication that the prompts influenced behaviour came from one practice, which improved the awareness of the prescribing clerk and enabled her involvement in the work. This practice was alone in enrolling an administrative member of staff in the intervention work over the full study period although others were involved in educational outreach or initial search generation.

Change as a result of the implementation package

Across the trials, achievement of the four primary outcome indicators appeared marginally better in practices that had received educational outreach meetings than in practices that had not (*Table 16*). Practices allocated to the diabetes control and anticoagulation for AF implementation packages that received feedback without educational meetings appeared to perform marginally worse than the control practices that had not received these packages.

There were signals from the process evaluation practices that staff had internalised the key clinical messages in risky prescribing and, to some extent, in anticoagulation for AF and felt more likely to act in accordance. However, there were also continuing barriers to acting on messages routinely. These included the relative infrequency of clinicians seeing new cases of AF, the relative complexity of managing AF, difficulties finding information in patient records and competing demands within consultations.

The anticoagulation and risky prescribing practices did systematically review all patients not achieving recommended care – although this was repeated more than once throughout the year in only one practice. The other three practices either completed a review once or were still in the process of completing it by the end of the intervention period. Some practices recognised the work as ongoing and discussed putting a further audit cycle or check on progress into their systems.

TABLE 16 Primary outcome achievement rate by uptake of outreach visit, with non-intervention group for comparison (WP5)^a

	Diabetes control (N = 40)	Risky prescribing (N = 40)	BP control (N = 32)	Anticoagulation in AF (N = 32)
Outreach visit received				
<i>n</i>	20	25	11	11
Mean (SD) (%)	25.36 (5.69)	5.53 (3.89)	53.30 (7.02)	76.98 (4.72)
No outreach visit				
<i>n</i>	20	15	21	21
Mean (SD) (%)	22.97 (6.73)	5.90 (2.06)	52.27 (7.13)	69.20 (9.95)
Non-intervention group achievement				
<i>n</i>	40	40	32	32
Mean (SD) (%)	24.35 (6.89)	7.14 (5.05)	51.17 (4.93)	75.20 (9.99)
a The outcome data presented in this table use the most recent primary outcome data available for each of the randomised practices.				

The diabetes and BP control practices tended not to engage in the systematic review of patients, unless the outreach support did substantial preparatory work for them. They struggled with capacity and felt as if they had already poured extensive resources into, and had developed particular working structures for, addressing these topics; there was a collective failure to understand how the package differed from usual ways of working. They were resistant to changing these structures or adding in extra resources. As they continued to work within structures, and using reminders, oriented towards achieving QOF – with some small adjustments that might improve QOF achievement – it seemed unlikely that the intervention would have a lasting effect in these practices.

Discussion

We had set out to devise an implementation package that could be adapted to target a range of high-impact indicators and be delivered within existing resources and ways of working in general practice. Our process evaluation offers three main explanations of why the implementation package was effective for one of four indicators. First, we observed losses in fidelity from delivery through to enactment by practice teams. Second, the type and scale of targeted behaviour changes varied by indicator, so that practice teams may have perceived and exerted greater control over risky prescribing. Third, we had sought to ensure that targeted indicators were well aligned with existing clinical priorities; this inadvertently resulted in the implementation package being insufficiently differentiated from normal activity.

We observed significant losses in the receipt of intervention components. The overwhelming majority of trial practices received quarterly feedback reports but we could not mandate actions beyond receipt. Reports may have been circulated to only those already involved in delivering care, limiting the opportunity to maximise capacity (e.g. by employing administrative support). Our observations of a selected group of practices suggested that practice managers did not share the feedback reports as widely or as frequently as we had intended. This was despite being engaged in the process evaluation and, therefore, being aware of being observed. Just over half of trial practices accepted the offer of an outreach visit, whereas only a minority took up further outreach visits and support. Waiting for an outreach visit is likely to have been perceived as a reason not to start work in response to feedback reports. Although the implementation package aimed to offer explicit advice and encourage action-planning, practices remained uncertain about expectations and roles (e.g. who should attend outreach visit meetings or act on patient searches). These were relatively observable fidelity losses; the absence of any effects for process outcomes in the trials strongly suggests that any remaining implementation package effects for three indicators (diabetes control, BP control and anticoagulation for AF) dissipated before or at the point of clinical care.

With regard to targeted behaviours, all signals from the process evaluation suggest that practices welcomed support to improve achievement for indicators; however, the complexity and amount of work required to achieve treatment goals was a barrier. For the anticoagulation and risky prescribing practices, the actions required were relatively straightforward, involving one member of staff and one change to care. Practice teams addressing risky prescribing may have been further motivated to change following observable improvements in feedback reports. For the diabetes and BP control practices, there was little clear guidance on who to involve and what they could do differently (rather than doing more) to get from the current position to the desired outcomes, and practices had historically invested considerable time and resources in these areas. Because the link between treatment goals and specific practitioner actions was relatively weak, practice staff sometimes became disillusioned or disengaged over time, as their achievement did not seem to improve despite consistent effort. This was apparent in the way that some staff started to criticise or withhold the reports in order to manage its negative effect on their professional identities, or express demoralisation given perceived gaps between effort and achievement. Negative effects of feedback may have been mitigated by outreach educational visits; practices receiving visits appeared to fare better than those that did not.

The implementation package was specifically designed to align with priorities and existing initiatives such as the QOF, an underlying assumption being that this would enhance motivation and engagement. However, this alignment seemed to have the opposite effect; although priorities were generally aligned, the additional work needed to improve, for example, diabetes or BP control could probably not be achieved within existing ways of working or resources. Furthermore, our suggestions that practices review their team roles and processes of care are likely to have fallen on 'stony ground.' In terms of NPT, an intervention needs to be considered to be aligned with the individual and collective priorities of the team, and sufficiently differentiated from current practice to ensure the value of engaging with it. Our intervention was aligned with priorities, but perhaps insufficiently differentiated from the ongoing work of QOF achievement, and, therefore, practices were reluctant to alter existing work patterns and resources. Furthermore, given that practices were already working to capacity, there was little organisational 'slack' and, therefore, little will to change.

We highlight three study limitations. First, fidelity data collection was limited by technical and pragmatic considerations; for instance, we were unable to track uses of the searches on practice computers. Second, in-depth work was carried out in only eight practices, which were recruited actively into the study. This might result in more motivated practices (especially as we had reimbursed them for participation), or those who are more used to research. This did not seem to be the case. The practices were not outliers in terms of achievement, and covered a range of CCG areas and patient list sizes. Whereas some practices had prior involvement with research, most of them did not. Third, our in-depth work across several practices was still conducted within constraints. The limited opportunities for observational work in the practices had an impact on relationship-building, which probably affected the observational data as the researcher remained a visitor. This was also indicative of overall engagement with the study and lent important contextual insights into the practices. We relied on practitioner self-report of whether or not and how they changed their work with patients, and they may have been more positive about the impact of the intervention than was warranted. It would not have been feasible to observe any clinical interactions with patients, where some of the work would have occurred.

Conclusion

Our process evaluation suggests three main explanations for the partial effect of the implementation package and why it was effective for one of four QIs. First, we observed fidelity losses from delivery through to enactment by practice teams. Second, the type and scale of targeted behaviour changes varied by QI, especially those related to perceived control and clarity of actions required to achieve outcomes. Third, the package often seemed indistinguishable from normal activity. We also identified unintended negative as well as positive consequences, such as when continuing feedback showing lack of progression towards goals demotivated teams. The future design of implementation strategies should learn from our successes and failures and ensure that interventions are sufficiently robust to survive fidelity losses, clarify required expectations and roles, and are targeted efficiently at key patient, team and organisational determinants of behaviour.

Patient and public involvement

Roles and experience within ASPIRE

Our PPI panel has demonstrated commitment to and engagement with ASPIRE throughout the programme's lifetime. The panel chairperson (MR) was a co-investigator and member of the Programme Management Team. The PPI panel met at quarterly intervals throughout the ASPIRE programme and played an integral role in the development and conduct of our research. The panel's contributions included:

- WP1. The consensus panel included patient representatives. All panel member ratings were weighted equally, meaning that patients' views helped to determine which indicators would be taken forward.

- WPs 2 and 4. We debated the relative merits of opt-in versus opt-out methods for recruiting general practices to share anonymised patient data and take part in the trials. The panel took the view that it would expect practices to share such data for research as a matter of routine and were keen to ensure that ASPIRE involved 'typical' practices as far as possible. We were therefore able to cite the panel's support in our application for ethical review and in our subsequent communications to CCGs and general practices. In addition, the panel provided helpful suggestions to support recruitment (e.g. recommending the use of recorded delivery when sending 'opt-out' invitations).
- WP3. Panel input was explicitly designed into the implementation package development. The panel provided feedback throughout the process, considered the acceptability of proposed interventions to patients and the potential contribution of patient-mediated delivery mechanisms. In addition, one panel member assisted with the procurement process through which we identified partners to deliver our outreach sessions.
- WP5. The panel provided feedback on process evaluation methods and interpretation of emerging data. Prior to the trial, panel members were interviewed by our process evaluation researcher to gather their perspectives on how the implementation package might work. Their responses fed into the production of 'logic models' (Figures 18 and 19) outlining hypothesised method(s) of action.
- More broadly, the panel assisted our ongoing networking and relationship-building efforts with CCGs and practices in the region. Panel members suggested and organised meetings at local general practices, including introductions from lead GPs outlining initiatives to improve patient interaction and health outcomes.
- Panel members attended and contributed to the three stakeholder engagement events held during the ASPIRE programme.

Subjectively, our panel members have indicated that they considered the PPI to be sufficiently planned and resourced, that their opinions and perspectives were respected, and that they were treated as equal partners in programme discussions.

Beyond ASPIRE, our PPI panel members have become involved in other research studies (e.g. as steering group members) and have contributed to other bids and funded research led by our group. We intend to build on this foundation to sustain and further develop our PPI partnership.

The role of patient and public involvement in implementation research

Patient and public involvement is generally an essential requirement for research funding. Distinctions can be drawn between clinical research, which generally focuses on patients, and implementation research, which generally focuses on health professionals' behaviour. There is uncertainty about the role of PPI in the latter field; therefore, we explored and defined the roles of PPI in implementation research to inform good practice guidance.

We used a structured consensus process with a panel comprising our nine active PPI members and two researcher members. We drew on available literature to identify 21 potential PPI roles in research. The panel rated their agreement with roles independently online in relation to both implementation and clinical research. Disagreements were discussed at a face-to-face meeting prior to a second online rating of all roles. Median scores were calculated and a final meeting held to review findings and consider recommendations.

Ten panellists completed the consensus process. For clinical research, there was strong support and consensus for the role of PPI throughout most of the research process. For implementation research, there were eight roles with consensus and strong support, seven roles with consensus but weaker support and six roles with no consensus. Those with strong support included prioritising and shaping research questions, and advising on supporting participant recruitment strategies. There were more disagreements relating to the roles of PPI in implementation research than in clinical research. PPI was generally rated as contributing less to the design and management of implementation research than clinical research.

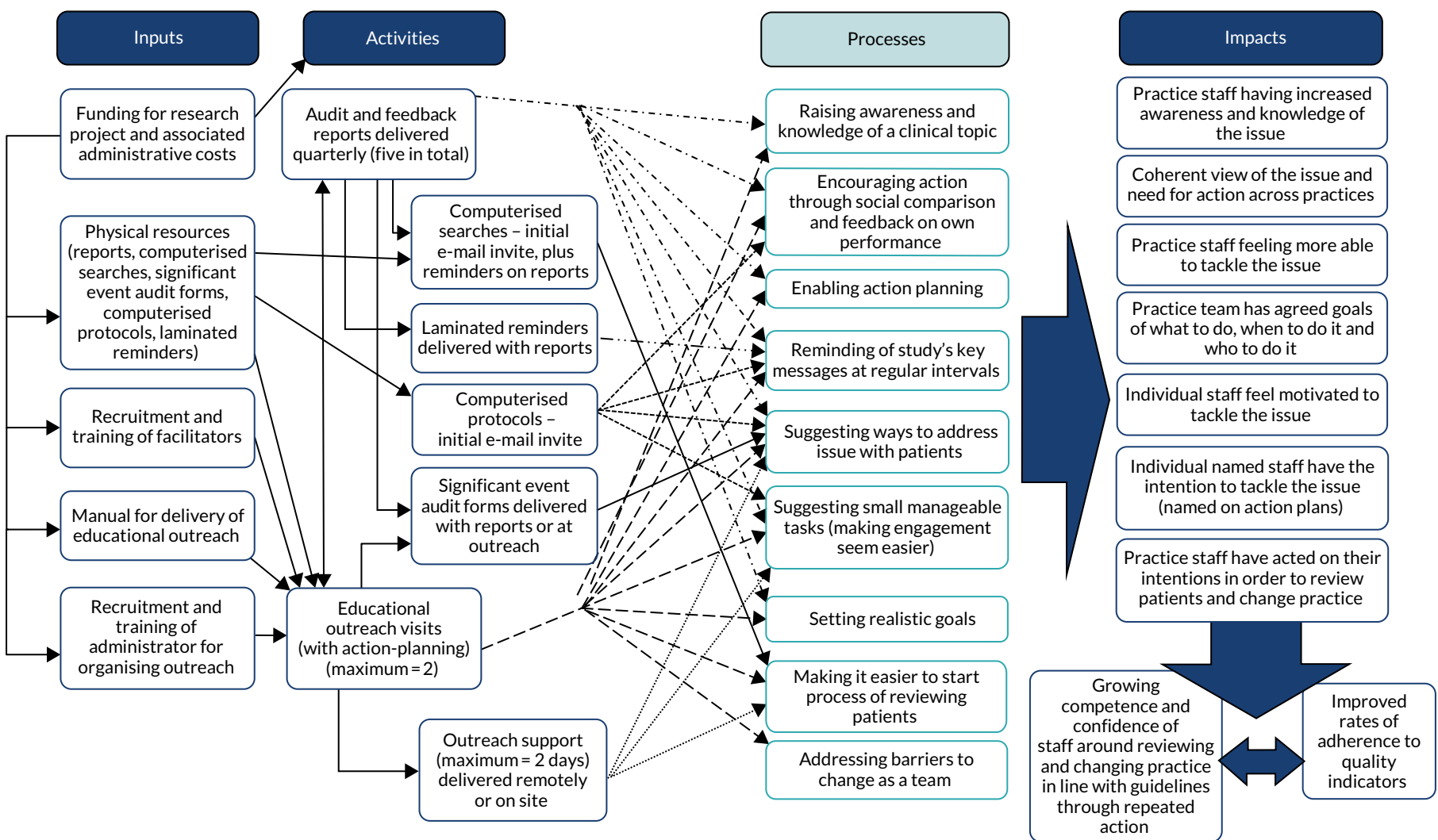


FIGURE 18 Process evaluation: intervention logic model (WP5).

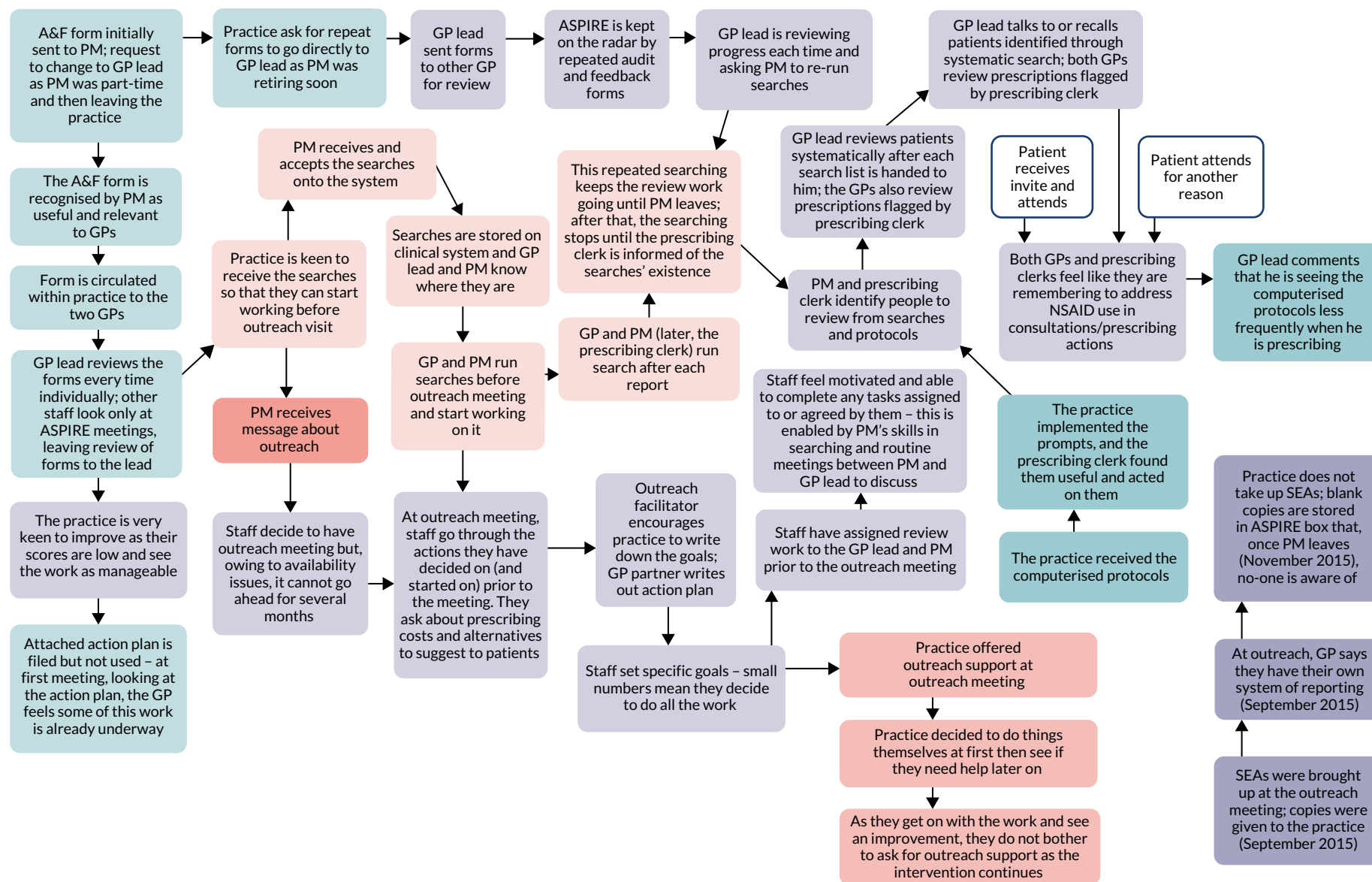


FIGURE 19 Process evaluation: example of implementation process model for Treetop Practice (WP5). PM, practice manager; SEA, significant event audit.

Patient and public involvement roles need to be tailored according to the nature of the research to ensure authentic and appropriate involvement. We provided a framework to guide the planning, conduct and reporting of PPI in implementation research, which has been published in Gray-Burrows *et al.*¹²²

Conclusions

We set out to design and apply an implementation package that could be delivered sustainably using resources typically available to primary care. We involved health professionals, commissioners and patients in structured deliberations to prioritise and develop a set of 'high-impact', evidence-based QIs associated with scope for improvement and that could be measured using routinely collected data. We demonstrated marked variations in indicator adherence among general practices and used a behavioural framework in interviews with health professionals to explore the reasons for variations in practice. We drew on these findings and existing evidence in work with stakeholders to develop a set of implementation packages adapted to target four indicators: avoidance of risky prescribing; treatment targets in type 2 diabetes; BP targets in treated hypertension; and anticoagulation in AF. The implementation packages were largely built around A&F, educational outreach and clinical prompts, within which we embedded BCTs and tailored content to each indicator.

Our pragmatic, cluster-randomised trials indicated the value of an implementation package targeting risky prescribing. However, the similarly adapted implementation packages targeting diabetes control, BP control and anticoagulation for AF were ineffective. Our mixed-method process evaluation highlighted cumulative losses in fidelity, especially around delivery and enactment. In the current pressurised context of primary care, practices found it difficult to achieve change for indicators related to behaviours perceived as less immediately within their control compared with relatively discrete prescribing decisions.

Particular programme strengths included a systematic approach to intervention development and evaluation, the use of opt-out practice recruitment and routinely collected data within a pair of 'real-world' balanced incomplete block cluster-randomised trials, and the participation of clinicians and patients throughout the programme. We also developed, with our PPI panel, a framework to guide the planning, conduct and reporting of PPI in implementation research.

Three features of our programme worked less well than we had hoped. First, we invested considerably in the 'diagnostic' phase of our programme. This limited the time and resources available to field test the implementation package to ensure that it could be delivered with reasonable fidelity. We would recommend dedicating less time to diagnostic work, especially as we already had a fair understanding of implementation challenges within the primary care context, and more time on iterative cycles of stress testing and enhancing the implementation packages before wider roll-out in the trials. Second, we focused on addressing determinants of practice largely based on the TDF and around evidence-based interventions that could feasibly be delivered within a cluster-randomised trial using general practices as the unit of allocation (e.g. A&F). Our ability to address some of the higher-level, meta-themes we identified in our interviews (WP2), such as internal and external sources of support, is more questionable. Such determinants are potentially less amenable to change and efforts to change them within our implementation packages were unsuccessful. We would encourage further work to develop and evaluate interventions targeting organisational and system-level influences on practice. Third, our 'opt-out' approach to recruit general practices to the randomised trials means that trial practices may not have engaged with the implementation package as much as if they had actively volunteered. However, an 'opt in' approach may also have limited the applicability of our findings to 'real-world' quality improvement initiatives targeting all practices within a given locality. We therefore suggest that others learn from our first lesson and ensure that interventions are designed to be sufficiently robust to withstand the buffeting they will inevitably receive in pragmatic trials in a challenging primary care context.

Publications lists and includes links to published articles resulting from this programme.

Implications for practice

For national agencies with implementation roles:

- There is a growing, if imperfect, evidence base on implementation strategies to promote the uptake of evidence-based practice,⁸³ as well as a variety of approaches to improvement underpinned by conviction more than evidence,¹²³ to consider in planning implementation.
- A realistic appraisal of the likelihood of being able to bring about change for a given targeted recommendation or behaviour can inform the selection of priorities for implementation support activities.

For CCGs, practice networks and general practices:

- We have demonstrated an approach to set priorities and to develop high-impact indicators that could be adapted for local quality improvement initiatives.
- Overambitious goals for change can undermine the motivation of general practice staff; focusing on continuous and cumulative improvements in evidence-based care has the potential to deliver major improvements in population health.
- The likelihood of successful improvement may depend on the degrees to which any goals for change are within the control of the professionals and patients who need to change behaviour.
- Gaps between evidence and practice are generally multifactorial and related to barriers operating at one or more of system, team, professional and patient levels.¹²⁴ We suggest that plans to tackle evidence–practice gaps take account of this and address barriers that are most amenable to change.
- Where there are existing routine data demonstrating variations in practice, A&F to support general practices, providing repeated feedback accompanied by persuasive messages, feasible goals and action plans has modest but scalable effects. Educational outreach visits may augment the delivery of feedback and help with action-planning. We demonstrated an approach to reduce risky prescribing that yielded an ICER of £2337 per incremental QALY, falling below the NICE preferred threshold of £20,000–30,000 cost per QALY.
- Computerised prompts can help change specific behaviours and are more likely to work if users need to provide a justification for over-riding recommendations;⁷⁰ but practices will circumvent them if they are too intrusive or disruptive.¹²⁵ Administrative staff processing repeat prescriptions can offer an alternative target to consultations.

For patient advocates and groups:

- Practice and commissioner patient groups face a wide-ranging quality agenda. This agenda could include taking an interest in variations between general practices in adherence to evidence-based practice. Supporting and making changes here could yield significant population benefits.
- It is important that participation in research to improve the quality of patient care is not dominated by general practices that are already more interested in research and quality. Patient groups may wish to encourage general practices to share data and take part in quality improvement studies.
- Our research suggests that a range of roles and expectations can be discussed with researchers when planning research to address professional and/or patient behaviour.

Implications for future research

For researchers:

- There are challenges in designing implementation strategies that are sufficiently robust to bring about change in the face of difficult clinical contexts and likely losses to fidelity. Maximise feasibility and ‘stress-testing’ work prior to rolling out interventions in a definitive, pragmatic trial.

- Our implementation package was unsuccessful for three QIs (type 2 diabetes control, BP control and use of anticoagulation for AF). Yet these are not intractable problems.⁶² Although implementation strategies may have small to modest effects, these can translate into worthwhile population health gains; therefore, we recommend further implementation research addressing these problems that builds on our lessons and the wider body of research literature. Concerted research strategies that target system, organisational and patient levels as well as general practices may be required to bring about more significant change.
- The ASPIRE programme effectively represented a nascent ‘implementation laboratory’ embedded within 10 CCGs.¹²⁶ It is possible to develop and test incremental ways of improving the delivery of health care that cumulatively both improve patient care and develop the scientific basis of health-care provision.¹²⁷ As well as cumulative improvement, implementation laboratories also offer a means of improving research efficiency and generalisability. Cluster-randomised trials typically require larger numbers of patients than individually randomised trials to account for lack of independence within clusters. Increasing the number of sites allows greater statistical efficiency than increasing the number of patients. Although this augments logistical challenges, embedding trials in an existing network or major improvement initiative facilitates recruitment and helps ensure ‘real-world’ generalisability. We recommend that researchers build collaborations with those responsible for large-scale regional or national improvement to establish implementation laboratories.
- Researchers will need to consider how best to recruit general practices to trials, given current research governance arrangements. We were able to recruit general practices to a pair of cluster-randomised trials using a relatively light-touch ‘opt-out’ approach. This both minimised any administrative burden for participating practices and ensured that our trials remained highly pragmatic. However, such an approach to this type of intervention study no longer appears feasible under research governance arrangements that require permission from general practices to participate in research as explicit research sites (a de facto opt-in approach).⁹⁴
- Considering PPI, distinctions can be drawn between clinical research, which generally focuses on patients, and implementation research, which generally focuses on health professionals’ behaviour. A flexible approach to PPI may help optimise the use of finite research and PPI resources. Researchers may find the framework we have developed to guide the planning, conduct and reporting of PPI in implementation research useful.¹²²

Ethics review

WPs 1, 2 and 3 were approved by the National Research Ethics Service Committee Yorkshire and the Humber – Leeds Central (12/YH/0254). WPs 4 and 5 were approved by the UK National Research Ethics Service (14/SC/1393). All WPs were granted local NHS research governance approval.

Acknowledgements

We thank all of the general practices that participated in this programme and our PPI panel. We also thank Wendy Hobson, Lauren Moreau and other staff from the University of Leeds for their administrative support. Thank you also to Elizabeth Andrews for her contribution to the development of WP1.

Contributions of authors

Robbie Foy (<https://orcid.org/0000-0003-0605-7713>) was principal investigator for the programme and led WP4a.

Thomas Willis (<https://orcid.org/0000-0002-0252-9923>) was the programme manager and led WP2.

Liz Glidewell (<https://orcid.org/0000-0003-2519-2654>) led WPs 3b and 5.

Rosie McEachan (<https://orcid.org/0000-0003-1302-6675>) led WP1.

Rebecca Lawton (<https://orcid.org/0000-0002-5832-402X>) led WP3a.

David Meads (<https://orcid.org/0000-0003-1369-2483>), **Daniele Bregantini** (<https://orcid.org/0000-0001-6802-1551>), **Laetitia Schmitt** (<https://orcid.org/0000-0003-1052-488X>) and **Armando Vargas-Palacios** (<https://orcid.org/0000-0002-6503-0134>) conducted the economic evaluation (WP4b), supervised by **Claire Hulme** (<https://orcid.org/0000-0003-2077-0419>).

Michelle Collinson (<https://orcid.org/0000-0003-3568-6455>), **Michael Holland** (<https://orcid.org/0000-0002-6331-5398>) and **Peter Heudtlass** (<https://orcid.org/0000-0002-9594-8050>) conducted trial sample calculations, randomisation, analysis and reporting, under the supervision of **Amanda Farrin** (<https://orcid.org/0000-0002-2876-0584>).

Cheryl Hunter (<https://orcid.org/0000-0002-1918-9662>) conducted the process evaluation, with guidance from **Vicky Ward** (<https://orcid.org/0000-0001-8684-0403>).

Robert West (<https://orcid.org/0000-0001-7305-3654>) contributed to the design of and conducted statistical analysis for WP2.

Suzanne Hartley (<https://orcid.org/0000-0003-2346-9461>) was responsible for operational delivery of the trials.

Paul Carder (<https://orcid.org/0000-0002-7940-6016>) and **Stella Johnson** (<https://orcid.org/0000-0002-2832-052X>) contributed to programme design and management, and advised on governance and data acquisition.

Sarah Alderson (<https://orcid.org/0000-0002-5418-0495>) provided clinical guidance, particularly during intervention design, and contributed to the economic evaluation.

Susan Clamp (<https://orcid.org/0000-0001-7994-3279>) contributed to intervention design.

Tim Stokes (<https://orcid.org/0000-0002-1127-1952>) advised on clinical recommendation selection and interpretation of WP2.

Gemma Louch (<https://orcid.org/0000-0001-6946-3693>), **Jane Heyhoe** (<https://orcid.org/0000-0002-9652-7928>) and **Emma Ingleson** (<https://orcid.org/0000-0001-6603-026X>) conducted interviews and analyses for WP3; Emma Ingleson also contributed towards intervention piloting in WP3b.

Martin Rathfelder (<https://orcid.org/0000-0002-1799-9987>) chaired the ASPIRE PPI panel and contributed to programme design.

Judith Richardson (<https://orcid.org/0000-0002-5129-813X>) contributed to programme design and advised on programme dissemination.

Bruno Rushforth (<https://orcid.org/0000-0001-7932-3619>) was involved in recommendation selection, design of data extraction for WP2 and led the reporting of WP1.

Duncan Petty (<https://orcid.org/0000-0003-2763-0726>) provided pharmacy advice, particularly during intervention design and trial delivery, managed the team of pharmacist facilitators involved in intervention delivery and contributed to the economic evaluation.

Ian Watt (<https://orcid.org/0000-0002-3147-8299>) contributed to programme design and advised on the selection of indicators for the trials.

Robbie Foy, Thomas Willis, Liz Glidewell, Rosie McEachan, Rebecca Lawton, Michelle Collinson and David Meads led the drafting of the report. All authors participated in programme management, WP management or dissemination meetings. All authors provided a critical review and final approval of the report.

Publications

Rushforth B, Stokes T, Andrews E, Willis TA, McEachan R, Faulkner S, Foy R. Developing 'high impact' guideline-based quality indicators for UK primary care: a multi-stage consensus process. *BMC Fam Prac* 2015;**16**:156.

Lawton R, Heyhoe J, Louch G, Ingleson E, Glidewell L, Willis TA, *et al*. Using the Theoretical Domains Framework (TDF) to understand adherence to multiple evidence-based indicators in primary care: a qualitative study. *Implement Sci* 2016;**11**:113.

Willis TA, Hartley S, Glidewell L, Farrin AJ, Lawton R, McEachan RRC, *et al*. Action to Support Practices Implement Research Evidence (ASPIRE): protocol for a cluster-randomised evaluation of adaptable implementation packages targeting 'high impact' clinical practice recommendations in general practice. *Implement Sci* 2016;**11**:1–11.

Glidewell L, Willis TA, Petty D, Lawton R, McEachan RRC, Ingleson E, *et al*. To what extent can behaviour change techniques be identified within an adaptable implementation package for primary care? A prospective directed content analysis. *Implement Sci* 2018;**13**:32.

Gray-Burrows KA, Willis TA, Foy R, Rathfelder M, Bland P, Chin A, *et al*. Role of patient and public involvement in implementation research: a consensus study *BMJ Qual Saf* 2018;**27**:858–864.

Willis TA, Collinson M, Glidewell L, Farrin AJ, Holland M, Meads D, *et al*. An adaptable implementation package targeting evidence-based indicators in primary care: a pragmatic cluster-randomised evaluation. *PLOS Med* 2020;**17**:e1003045.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. Rushforth B, Stokes T, Andrews E, Willis TA, McEachan R, Faulkner S, Foy R. Developing 'high impact' guideline-based quality indicators for UK primary care: a multi-stage consensus process. *BMC Fam Pract* 2015;**16**:156. <https://doi.org/10.1186/s12875-015-0350-6>
2. Eccles M, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. *Implement Sci* 2009;**4**:18. <https://doi.org/10.1186/1748-5908-4-18>
3. Cooksey D. *A Review of UK Health Research Funding*. HM Treasury on behalf of the Controller of Her Majesty's Stationery Office; 2006.
4. Dawda P, Jenkins R, Varnam R. *Quality Improvement in General Practice*. London: The King's Fund; 2010.
5. Duerden M, Millson D, Avery A, Smart S. *The Quality of GP Prescribing*. London: The King's Fund; 2011.
6. Smith J, Holder H, Edwards N, Maybin J, Parker H, Rosen R, et al. *Securing the Future of General Practice. New Models of Primary Care*. London: Nuffield Trust; 2013.
7. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, Salisbury C, National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet* 2016;**387**:2323–30. [https://doi.org/10.1016/S0140-6736\(16\)00620-6](https://doi.org/10.1016/S0140-6736(16)00620-6)
8. Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. *Qual Health Care* 2001;**10**:152–8. <https://doi.org/10.1136/qhc.0100152>
9. Appleby J, Ham C, Imison C, Jennings M. *Improving NHS Productivity: More with the Same Not More of the Same*. London: The King's Fund; 2010.
10. Roland M. Linking physicians' pay to the quality of care – a major experiment in the United Kingdom. *N Engl J Med* 2004;**351**:1448–54. <https://doi.org/10.1056/NEJMp041294>
11. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003;**15**:523–30. <https://doi.org/10.1093/intqhc/mzg081>
12. Hutchinson A, McIntosh A, Anderson J, Gilbert C, Field R. Developing primary care review criteria from evidence-based guidelines: coronary heart disease as a model. *Br J Gen Pract* 2003;**53**:690–6.
13. Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators – a systematic review. *Implement Sci* 2012;**7**:21. <https://doi.org/10.1186/1748-5908-7-21>
14. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;**2**(3). <https://doi.org/10.3310/hta2030>
15. Stelfox HT, Straus SE. Measuring quality of care: considering measurement frameworks and needs assessment to guide quality indicator development. *J Clin Epidemiol* 2013;**66**:1320–7. <https://doi.org/10.1016/j.jclinepi.2013.05.018>
16. Campbell SM, Hann M, Hacker J, Durie A, Thapar A, Roland MO. Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. *Qual Saf Health Care* 2002;**11**:125–30. <https://doi.org/10.1136/qhc.11.2.125>

17. NHS Employers. *Summary of Changes to QOF 2014/15 – England*. 2014. URL: www.england.nhs.uk/wp-content/uploads/2014/05/gms-contract-04-14.pdf (accessed 20 February 2020).
18. Chassin MR, Loeb JM, Schmaltz SP, Wachter RM. Accountability measures – using measurement to promote quality improvement. *N Engl J Med* 2010;**363**:683–8. <https://doi.org/10.1056/NEJMs1002320>
19. Guthrie B, Inkster M, Fahey T. Tackling therapeutic inertia: role of treatment data in quality indicators. *BMJ* 2007;**335**:542–4. <https://doi.org/10.1136/bmj.39259.400069.AD>
20. Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, Lilford RJ. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care* 2008;**17**:170–7. <https://doi.org/10.1136/qshc.2007.023655>
21. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;**13**:475–80. <https://doi.org/10.1093/intqhc/13.6.475>
22. Baker R, England J. Should we use outcomes data to help manage general practice? *Br J Gen Pract* 2014;**64**:e804–6. <https://doi.org/10.3399/bjgp14X683005>
23. NHS Digital. *Diabetes Audit 2011–12 Report 1: Care Processes and Treatment Targets*. Leeds: NHS Digital; 2013.
24. National Institute for Health and Care Excellence (NICE). *MI: Secondary Prevention*. Secondary Prevention in Primary and Secondary Care for Patients Following a Myocardial Infarction (CG48). London: NICE; 2007.
25. Dreischulte T, Grant A, Donnan P, McCowan C, Davey P, Petrie D, et al. A cluster randomised stepped wedge trial to evaluate the effectiveness of a multifaceted information technology-based intervention in reducing high-risk prescribing of non-steroidal anti-inflammatory drugs and antiplatelets in primary medical care: the DQIP study protocol. *Implement Sci* 2012;**7**:24. <https://doi.org/10.1186/1748-5908-7-24>
26. Steel N, Abdelhamid A, Stokes T, Edwards H, Fleetcroft R, Howe A, Qureshi N. A review of clinical practice guidelines found that they were often based on evidence of uncertain relevance to primary care patients. *J Clin Epidemiol* 2014;**67**:1251–7. <https://doi.org/10.1016/j.jclinepi.2014.05.020>
27. Agency for Healthcare Research and Quality. *The Ambulatory Care Quality Alliance Recommended Starter Set*. Agency for Healthcare Research and Quality. 2005. URL: <http://archive.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/ambulatory-care/starter-set.html> (accessed 2 March 2015).
28. Institute of Medicine, Board on Health Care Services, Committee on Redesigning Health Insurance Performance Measures, Payment and Performance Improvement Programs. *Performance Measurement: Accelerating improvement*. Washington, DC: The National Academies Press; 2006. <https://doi.org/10.17226/11517>
29. Lambert MF, Cook JV, Roelant E, Bradshaw C, Foy R, Eccles MP. Feasibility of applying review criteria for depression and osteoporosis national guidance in primary care. *Prim Health Care Res Dev* 2014;**15**:396–405. <https://doi.org/10.1017/S146342361400005X>
30. Loxterkamp D. Humanism in the time of metrics – an essay by David Loxterkamp. *BMJ* 2013;**347**:f5539. <https://doi.org/10.1136/bmj.f5539>
31. Shekelle PG. Quality indicators and performance measures: methods for development need more standardization. *J Clin Epidemiol* 2013;**66**:1338–9. <https://doi.org/10.1016/j.jclinepi.2013.06.012>
32. Campbell SM, Kontopantelis E, Hannon K, Burke M, Barber A, Lester HE. Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. *BMC Fam Pract* 2011;**12**:85. <https://doi.org/10.1186/1471-2296-12-85>

33. Willis TA, West R, Rushforth B, Stokes T, Glidewell L, Carder P, *et al.* Variations in achievement of evidence-based, high-impact quality indicators in general practice: an observational study. *PLOS ONE* 2017;**12**:e0177949. <https://doi.org/10.1371/journal.pone.0177949>
34. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;**348**:2635–45. <https://doi.org/10.1056/NEJMs022615>
35. Baker R, Roland M. General practice: continuous quality improvement since 1948. *Br J Gen Pract* 2002;**52**:S2–3.
36. Campbell SM, Roland MO, Middleton E, Reeves D. Improvements in quality of clinical care in English general practice 1998-2003: longitudinal observational study. *BMJ* 2005;**331**:1121. <https://doi.org/10.1136/bmj.38632.611123.AE>
37. Kirk SA, Campbell SM, Kennell-Webb S, Reeves D, Roland MO, Marshall MN. Assessing the quality of care of multiple conditions in general practice: practical and methodological problems. *Qual Saf Health Care* 2003;**12**:421–7. <https://doi.org/10.1136/qhc.12.6.421>
38. Steel N, Bachmann M, Maisey S, Shekelle P, Breeze E, Marmot M, Melzer D. Self reported receipt of care consistent with 32 quality indicators: national population survey of adults aged 50 or more in England. *BMJ* 2008;**337**:a957. <https://doi.org/10.1136/bmj.a957>
39. Abdelhamid AS, Maisey S, Steel N. Predictors of the quality of care for asthma in general practice: an observational study. *Fam Pract* 2010;**27**:186–91. <https://doi.org/10.1093/fampra/cmp095>
40. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;**309**:1351–2. <https://doi.org/10.1001/jama.2013.393>
41. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;**2**:3. <https://doi.org/10.1186/2047-2501-2-3>
42. NHS Right Care. *NHS Atlas of Variation in Healthcare*. London: Public Health England; 2015.
43. Department of Health and Social Care (DHSC). *Cardiovascular Disease Outcomes Strategy. Improving Outcomes for People with or at Risk of Cardiovascular Disease*. London: DHSC; 2013.
44. Lord PA, Willis TA, Carder P, West RM, Foy R. Optimizing primary care research participation: a comparison of three recruitment methods in data-sharing studies. *Fam Pract* 2016;**33**:200–4. <https://doi.org/10.1093/fampra/cmw003>
45. Public Health England. *National General Practice Profiles*. URL: <http://fingertips.phe.org.uk/profile/general-practice> (accessed 14 March 2014).
46. Cohen J. A power primer. *Psychol Bull* 1992;**112**:155–9. <https://doi.org/10.1037//0033-2909.112.1.155>
47. Prescribing and Primary Care Team – Health and Social Care Information Centre (HSCIC). *Quality and Outcomes Framework Achievement, Prevalence and Exceptions Data, 2012/13*. Leeds: HSCIC; 2013. <https://doi.org/10.7748/phc.22.3.13.s10>
48. Down L, Metcalfe C, Avery K, Noble S, Lane JA, Neal DE, *et al.* Factors distinguishing general practitioners who more readily participated in a large randomized trial were identified. *J Clin Epidemiol* 2009;**62**:67–73. <https://doi.org/10.1016/j.jclinepi.2008.02.014>
49. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care* 2007;**45**:489–96. <https://doi.org/10.1097/MLR.0b013e31803bb479>

50. Levine DM, Linder JA, Landon BE. The quality of outpatient care delivered to adults in the United States, 2002 to 2013. *JAMA Intern Med* 2016;**176**:1778–90. <https://doi.org/10.1001/jamainternmed.2016.6217>
51. Rushforth B, McCrorie C, Glidewell L, Midgley E, Foy R. Barriers to effective management of type 2 diabetes in primary care: qualitative systematic review. *Br J Gen Pract* 2016;**66**:e114–27. <https://doi.org/10.3399/bjgp16X683509>
52. Lawton R, Heyhoe J, Louch G, Ingleson E, Glidewell L, Willis TA, et al. Using the Theoretical Domains Framework (TDF) to understand adherence to multiple evidence-based indicators in primary care: a qualitative study. *Implement Sci* 2016;**11**:113. <https://doi.org/10.1186/s13012-016-0479-2>
53. Grol RP, Bosch MC, Hulscher ME, Eccles MP, Wensing M. Planning and studying improvement in patient care: the use of theoretical perspectives. *Milbank Q* 2007;**85**:93–138. <https://doi.org/10.1111/j.1468-0009.2007.00478.x>
54. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A, 'Psychological Theory' Group. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005;**14**:26–33. <https://doi.org/10.1136/qshc.2004.011155>
55. Dreischulte T, Grant AM, McCowan C, McAnaw JJ, Guthrie B. Quality and safety of medication use in primary care: consensus validation of a new set of explicit medication assessment criteria and prioritisation of topics for improvement. *BMC Clin Pharmacol* 2012;**12**:5. <https://doi.org/10.1186/1472-6904-12-5>
56. National Institute for Health and Care Excellence (NICE). *Type 2 Diabetes: The Management of Type 2 Diabetes (NCG 87)*. London: NICE; 2009.
57. Krause T, Lovibond K, Caulfield M, McCormack T, Williams B, Guideline Development Group. Management of hypertension: summary of NICE guidance. *BMJ* 2011;**343**:d4891. <https://doi.org/10.1136/bmj.d4891>
58. Jones C, Pollit V, Fitzmaurice D, Cowan C, Guideline Development Group. The management of atrial fibrillation: summary of updated NICE guidance. *BMJ* 2014;**348**:g3655. <https://doi.org/10.1136/bmj.g3655>
59. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ* 2000;**320**:50–2. <https://doi.org/10.1136/bmj.320.7226.50>
60. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ* 2004;**328**:343–5. <https://doi.org/10.1136/bmj.328.7435.343>
61. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ* 2007;**334**:455–9. <https://doi.org/10.1136/bmj.39108.379965.BE>
62. Shojania KG, Ranji SR, McDonald KM, Grimshaw JM, Sundaram V, Rushakoff RJ, Owens DK. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006;**296**:427–40. <https://doi.org/10.1001/jama.296.4.427>
63. Baiardini I, Braidò F, Bonini M, Compalati E, Canonica GW. Why do doctors and patients not follow guidelines? *Curr Opin Allergy Clin Immunol* 2009;**9**:228–33. <https://doi.org/10.1097/ACI.0b013e32832b4651>
64. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implement Sci* 2009;**4**:54. <https://doi.org/10.1186/1748-5908-4-54>

65. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, *et al.* Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;**348**:g1687. <https://doi.org/10.1136/bmj.g1687>
66. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, *et al.* The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013;**46**:81–95. <https://doi.org/10.1007/s12160-013-9486-6>
67. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, *et al.* Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;**6**:CD000259. <https://doi.org/10.1002/14651858.CD000259.pub3>
68. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, *et al.* Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007;**4**:CD000409. <https://doi.org/10.1002/14651858.CD000409.pub2>
69. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, *et al.* Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *BMJ* 2013;**346**:f657. <https://doi.org/10.1136/bmj.f657>
70. Shojania KG, Jennings A, Mayhew A, Ramsay CR, Eccles MP, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. *Cochrane Database Syst Rev* 2009;**3**:CD001096. <https://doi.org/10.1002/14651858.CD001096.pub2>
71. Glidewell L, Willis TA, Petty D, Lawton R, McEachan RRC, Ingleson E, *et al.* To what extent can behaviour change techniques be identified within an adaptable implementation package for primary care? A prospective directed content analysis. *Implement Sci* 2018;**13**:32. <https://doi.org/10.1186/s13012-017-0704-7>
72. Primary Care Commissioning CIC. QOF Business Rules v23.0. URL: www.pcc-cic.org.uk/article/qof-business-rules-v230 (accessed 2 March 2015).
73. Willis TA, Hartley S, Glidewell L, Farrin AJ, Lawton R, McEachan RR, *et al.* Action to Support Practices Implement Research Evidence (ASPIRE): protocol for a cluster-randomised evaluation of adaptable implementation packages targeting 'high impact' clinical practice recommendations in general practice. *Implement Sci* 2016;**11**:25. <https://doi.org/10.1186/s13012-016-0387-5>
74. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;**350**:h2147. <https://doi.org/10.1136/bmj.h2147>
75. Office for National Statistics. *Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2014*. 2015. URL: www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tc%3A77-368259 (accessed 26 January 2016).
76. Howard RL, Avery AJ, Slavenburg S, Royal S, Pipe G, Lucassen P, Pirmohamed M. Which drugs cause preventable admissions to hospital? A systematic review. *Br J Clin Pharmacol* 2007;**63**:136–47. <https://doi.org/10.1111/j.1365-2125.2006.02698.x>
77. Aguilar MI, Hart R. Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database Syst Rev* 2005;**3**:CD001927. <https://doi.org/10.1002/14651858.CD001927.pub2>
78. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, *et al.* Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *Lancet* 2014;**383**:955–62. [https://doi.org/10.1016/S0140-6736\(13\)62343-0](https://doi.org/10.1016/S0140-6736(13)62343-0)

79. Sundstrom J, Arima H, Woodward M, Jackson R, Karmali K, Lloyd-Jones D, *et al.* Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet* 2014;**384**:591–8. [https://doi.org/10.1016/S0140-6736\(14\)62070-5](https://doi.org/10.1016/S0140-6736(14)62070-5)
80. Hartley S, Carder P, Foy R, Heudtlass P, Glidewell L, Ingleson E, *et al.* Optimising participation and generalisability: the use of opt-out recruitment for an implementation trial in primary care. *Trials* 2015;**16**:076. <https://doi.org/10.1186/1745-6215-16-S2-O76>
81. Heudtlass P, Foy R, Glidewell L, Hartley S, Turgoose J, Willis T, *et al.* Use of electronic health records in the design and evaluation of implementation research in primary care - experiences from the aspire cluster randomised controlled trials (CRCTS). *Trials* 2015;**16**:P33. <https://doi.org/10.1186/1745-6215-16-S2-P33>
82. National Institute for Health and Care Excellence (NICE). *Atrial fibrillation: The Management of Atrial Fibrillation* (CG180). London: NICE; 2014. URL: www.nice.org.uk/guidance/cg180?UNLID=81235962120166782425 (accessed 28 February 2019).
83. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. *Implement Sci* 2012;**7**:50. <https://doi.org/10.1186/1748-5908-7-50>
84. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al.* Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;**8**(6). <https://doi.org/10.3310/hta8060>
85. NHS Digital. *General and Personal Medical Services, England, 2005–2015, as at 30 September, Provisional Experimental Statistics*. 2016. URL: <https://digital.nhs.uk/catalogue/PUB20503> (accessed 28 February 2019).
86. Health Education England. *Health Education England: Working Across Yorkshire and the Humber*. URL: <https://hee.nhs.uk/hee-your-area/yorkshire-humber> (accessed 15 April 2017).
87. Willis TA, Collinson M, Glidewell L, Farrin AJ, Holland M, Meads D, *et al.* An adaptable implementation package targeting evidence-based indicators in primary care: a pragmatic cluster-randomised evaluation. *PLOS Med* 2020;**17**:e1003045. <https://doi.org/10.1371/journal.pmed.1003045>
88. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer prescribing – a trial of education, informatics, and financial incentives. *N Engl J Med* 2016;**374**:1053–64. <https://doi.org/10.1056/NEJMs1508955>
89. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, *et al.* Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet* 2012;**379**:2252–61. [https://doi.org/10.1016/S0140-6736\(12\)60480-2](https://doi.org/10.1016/S0140-6736(12)60480-2)
90. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014;**67**:267–77. <https://doi.org/10.1016/j.jclinepi.2013.08.015>
91. Berwick DM. The science of improvement. *JAMA* 2008;**299**:1182–4. <https://doi.org/10.1001/jama.299.10.1182>
92. Davies HTO, Nutley SM, Smith PC. *What works? Evidence-Based Policy and Practice in Public Services*. Bristol; Policy Press; 2000. <https://doi.org/10.2307/j.ctt1t892t3>
93. Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989;**67**:268–317. <https://doi.org/10.2307/3350142>
94. Mills P. *HRA Approval in Primary Care Settings: Principles of Study Set-up*. London: NHS Health Research Authority; 2017.

95. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655. <https://doi.org/10.1136/bmj.a1655>
96. Hallsworth M, Chadborn T, Sallis A, Sanders M, Berry D, Greaves F, *et al.* Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *Lancet* 2016;**387**:1743–52. [https://doi.org/10.1016/S0140-6736\(16\)00215-4](https://doi.org/10.1016/S0140-6736(16)00215-4)
97. Foy R, Leaman B, McCrorie C, Petty D, House A, Bennett M, *et al.* Prescribed opioids in primary care: cross-sectional and longitudinal analyses of influence of patient and practice characteristics. *BMJ Open* 2016;**6**:e010276. <https://doi.org/10.1136/bmjopen-2015-010276>
98. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia* 2013;**56**:1925–33. <https://doi.org/10.1007/s00125-013-2940-y>
99. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal (PMG9)*. London: NICE; 2013.
100. Department of Health and Social Care (DHSC). *NHS Reference Costs 2015 to 2016*. London: DHSC; 2016.
101. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: PSSRU; 2016.
102. The Campbell and Cochrane Economics Methods Group (CCEMG), Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre). CCEMG – EPPI-Centre Cost Converter. 2016. URL: <http://eppi.ioe.ac.uk/costconversion/default.aspx> (accessed 28 February 2019).
103. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions. *Health Econ* 2004;**13**:405–15. <https://doi.org/10.1002/hec.903>
104. National Institute for Health and Care Excellence (NICE). *Osteoarthritis: Care and Management (CG177)*. London; 2014.
105. National Centre for Social Research. *Health Survey for England, 1998. 2000*. URL: <https://discover.ukdataservice.ac.uk/catalogue/?sn=4150> (accessed 18 August 2017).
106. National Institute for Health and Care Excellence (NICE). *Acute Kidney Injury: Prevention, Detection and Management (CG169)*. London: NICE; 2013.
107. Lovibond K, Jowett S, Barton P, Caulfield M, Heneghan C, Hobbs FD, *et al.* Cost-effectiveness of options for the diagnosis of high blood pressure in primary care: a modelling study. *Lancet* 2011;**378**:1219–30. [https://doi.org/10.1016/S0140-6736\(11\)61184-7](https://doi.org/10.1016/S0140-6736(11)61184-7)
108. National Institute for Health and Care Excellence (NICE). *Hypertension in Adults: Diagnosis and Management (CG127)*. London; NICE; 2016.
109. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular-disease risk profiles. *Am Heart J* 1991;**121**:293–8. [https://doi.org/10.1016/0002-8703\(91\)90861-B](https://doi.org/10.1016/0002-8703(91)90861-B)
110. Payne R. *Excel Worksheet for Calculating Cardiovascular Risk*. URL: <http://cvrisk.mvm.ed.ac.uk/calculator/excelcalc.htm> (accessed 28 February 2018).
111. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *BMJ* 2010;**341**:c6624. <https://doi.org/10.1136/bmj.c6624>
112. Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al.* A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess* 2007;**11**(14). <https://doi.org/10.3310/hta11140>

113. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ* 2009;**338**:b1665. <https://doi.org/10.1136/bmj.b1665>
114. National Institute for Health and Care Excellence (NICE). *Type 2 Diabetes in Adults: Management* (NG28). London: NICE; 2017.
115. Hoomans T, Fenwick EA, Palmer S, Claxton K. Value of information and value of implementation: application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value Health* 2009;**12**:315–24. <https://doi.org/10.1111/j.1524-4733.2008.00431.x>
116. Hoomans T, Severens JL. Economic evaluation of implementation strategies in health care. *Implement Sci* 2014;**9**:168. <https://doi.org/10.1186/s13012-014-0168-y>
117. Walker S, Dixon S, Palmer S, Sculpher M. *Getting cost-effectiveness technologies into practice: the value of implementation*. Research Report 014. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York; 2013.
118. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, *et al*. Understanding the implementation of complex interventions in health care: the normalization process model. *BMC Health Serv Res* 2007;**7**:148. <https://doi.org/10.1186/1472-6963-7-148>
119. May CR, Mair FS, Dowrick CF, Finch TL. Process evaluation for complex interventions in primary care: understanding trials using the normalization process model. *BMC Fam Pract* 2007;**8**:42. <https://doi.org/10.1186/1471-2296-8-42>
120. Ritchie J, Spencer L. Qualitative Data Analysis for Applied Policy Research. In Bryman A, Burgess RG, editors. *Analysing Qualitative Data*. London: Routledge; 1994. pp. 173–94.
121. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;**13**:117. <https://doi.org/10.1186/1471-2288-13-117>
122. Gray-Burrows KA, Willis TA, Foy R, Rathfelder M, Bland P, Chin A, *et al*. Role of patient and public involvement in implementation research: a consensus study. *BMJ Qual Saf* 2018;**27**:858–64. <https://doi.org/10.1136/bmjqs-2017-006954>
123. Grol R. Personal paper. Beliefs and evidence in changing clinical practice. *BMJ* 1997;**315**:418–21. <https://doi.org/10.1136/bmj.315.7105.418>
124. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *Milbank Q* 2001;**79**:281–315. <https://doi.org/10.1111/1468-0009.00206>
125. Rousseau N, McColl E, Newton J, Grimshaw J, Eccles M. Practice based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care. *BMJ* 2003;**326**:314. <https://doi.org/10.1136/bmj.326.7384.314>
126. Ivers NM, Grimshaw JM. Reducing research waste with implementation laboratories. *Lancet* 2016;**388**:547–8. [https://doi.org/10.1016/S0140-6736\(16\)31256-9](https://doi.org/10.1016/S0140-6736(16)31256-9)
127. Halpern D, Mason D. Radical incrementalism. *Evaluation* 2015;**21**:143–9. <https://doi.org/10.1177/1356389015578895>

Appendix 1 Supplementary tables and figures

TABLE 17 Description of developed implementation package following the template for intervention description and replication (TIDieR) checklist⁶⁵ (WP3b)

	A&F	Educational outreach (supplemented by A&F)	Computerised and manual prompts and reminders
Rationale	To develop an implementation package that can be applied within existing primary care systems and resources and adapted to target specific determinants of change for four QIs		
Control interventions	Both control and intervention practices were exposed to standard 'background' practice quality improvement initiatives (e.g. national guidelines, QOF)		
Materials and training	<p>Practice-specific quarterly audit reports. Each report compared practice achievement of targeted indicators with those of other participating practices within their CCG, with all participating practices and trends over time. Reports encouraged reflection on progress and prompted change. Reports included information on clinical recommendations, suggested change strategies and suggested consequences of inaction. Practices were encouraged to set goals based on graded tasks (based on the number of clinical recommendations and the number of patients requiring attention) and use an action-planning template specifying who would do what, in what circumstances, and how and when the achievement would be reviewed. Subsequent reports included potential actions identified during outreach sessions</p> <p>Computerised searches were offered to systematically identify all patients whose care required review and to facilitate repeat searching</p> <p>Short and longer significant event audit templates were developed for risky prescribing and anticoagulation for AF indicators to facilitate root cause analyses and action-planning from harmful events or near misses</p>	<p>We commissioned and recruited experienced pharmacists as facilitators. The pharmacists received 2 days' training aimed to increase motivation, prompt individual and group reflection, increase confidence and intention to act. For each outreach visit, a practice-specific outreach pack was developed containing the most recent (and all previous) audit report(s); a session outline; an action plan template that included space for noting current performance, setting a target, identifying who will do what and review date; and templates for assessing costs and benefits</p>	<p>For risky prescribing nine computerised prompts were developed to be triggered within the consultation and during repeat prescribing on the basis of a clinical code algorithm for age/diagnosis and drug. When triggered a brief message notified that the patient was at risk and presented an evidence-based statement (e.g. 'This patient has CKD. NSAID use accounts for an estimated 15% of all cases of acute renal failure and 36% of drug-induced cases'). A one-click justification was required (e.g. continue with risk, add medication or stop medication)</p> <p>Two prompts were developed for anticoagulation for AF but could not be made available within the study timelines</p> <p>Patient-directed checklists were developed to facilitate shared decision-making for managing BP and diabetes outcomes but could not be made available within the study timelines</p> <p>Reminders in the form of laminated information sheets were created to convey key clinical information (BP and anticoagulation for AF)</p> <p>Pens and sticky notes were sent to all practices with a topic-specific reminder to prompt behaviour</p>
			continued

TABLE 17 Description of developed implementation package following the template for intervention description and replication (TIDieR) checklist⁶⁵ (WP3b) (*continued*)

	A&F	Educational outreach (supplemented by A&F)	Computerised and manual prompts and reminders
Supportive activities	None	Two-day pharmacist training included a one-day face-to-face meeting with intervention developers focusing on goal-setting, action-planning, clinical barriers and persuasive communication. This was followed by a half day of independent study using a folder of supporting documentation relevant to each indicator. The first outreach meeting of each facilitator was observed by an experienced facilitator and feedback given	None
Intervention provider	The research team	Trained pharmacists	The research team
Mode of delivery	<p>Reports were sent by post and e-mail. Practices were sent invitations to use computerised searches via their clinical information system</p> <p>Researchers e-mailed the practice manager and colleagues introducing significant event audit templates</p>	Face-to-face sessions were offered to practices	Practices were invited to use computerised prompts via their clinical information system. Researchers e-mailed the practice manager and colleagues recommending local activation of the prompts
Schedule and intensity	Quarterly feedback reports. Practices were offered access to searches and significant event audit templates at the beginning of the study and reminded of their availability via quarterly feedback reports	Practices were offered an initial 30-minute session. All relevant practice staff were invited to identify and review patients requiring action. A key clinical contact was identified to support practice engagement. Initial visits focused on practice achievement data (from audit reports), identifying models of good practice, addressing barriers to change and creating an action plan to facilitate and review change. Up to 2 days of pharmacist time were offered to support patient identification and review. An additional follow-up visit was offered to review progress and support the practice to create more challenging or attainable action plans	Practices were offered access to prompts at the beginning of the study and reminded of their availability via quarterly feedback reports. Practices were offered access to checklists at the beginning of the study and reminded of their availability via quarterly feedback reports. Sticky notes and pens were sent to all practices
Tailoring	Searches could be tailored by practices, allowing them to identify patients relevant to all or individual recommendations, or adjust target values to select specific groups of patients	Session content could be modified to practice requirements	Prompts could be copied and modified to practice requirements
Modifications	None		
Fidelity of delivery, receipt and enactment	Assessed in the subsequent process evaluation		

BOX 2 Illustrative denominator and numerator information for two 'high-impact' QIs (WP1)

Example 1: secondary prevention of MI**Recommendation**

All patients who have had an acute MI should be offered treatment with a combination of the following drugs: angiotensin-converting-enzyme inhibitor; aspirin; beta-blocker; statin.

Statements linked by logical operators to describe numerators and denominators**Denominator**

Coding for MI (drawing on QOF Read code clusters used by QOF indicator CHD14 – although note difference in our requirement of any prior MI coding).

Numerator

Coding for MI (drawing on QOF Read code clusters used by QOF indicator CHD14 – although note difference in our requirement of any prior myocardial infarction coding) AND [currently treated with: ACE inhibitor OR ARB (angiotensin II receptor blocker)] AND (currently treated with: aspirin OR alternative anti-platelet medication) AND (currently treated with a beta-blocker) AND (currently treated with a statin).

Example 2: BP targets for hypertension**Composite recommendation**

- a. Aim for a target clinic BP < 140/90 mmHg in people aged under 80 years with treated hypertension.
- b. Aim for a target clinic BP < 150/90 mmHg in people aged 80 years and over with treated hypertension.

Statements linked by logical operators to describe numerators and denominators**Denominators**

- a. On hypertension register AND age < 80 years (drawing on QOF Read code clusters used by QOF indicator BP5 – although note additional age criteria).
- b. On hypertension register AND age ≥ 80 years (drawing on QOF Read code clusters used by QOF indicator BP5 – although note additional age criteria).

Numerators

- a. On hypertension register AND age < 80 years AND last BP (measured in the preceding 9 months) < 140/90 mmHg (drawing on QOF Read code clusters used by QOF indicator BP5 – although note additional age criteria and lower BP target compared to QOF).
- b. On hypertension register AND age ≥ 80 years AND last BP (measured in the preceding 9 months) < 150/90 mmHg (drawing on QOF Read code clusters used by QOF indicator BP5 – although note additional age criteria).

Reproduced from Rushforth *et al.*¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

TABLE 18 Behaviour change techniques excluded from intervention development or intended but not subsequently operationalised (WP3b)

BCTs for changing determinants of behaviour ⁵²		BCTs excluded because of delivery mechanism or contextual constraints (BCT taxonomy code reference) ⁶⁶	BCTs intended but not operationalised (BCT taxonomy code reference) ⁶⁶
<i>Relevant determinants</i>			
Core determinants 'environmental context' and 'social and professional role'	Social support	Social support emotional (3.3)	
	Antecedents	Avoidance/reducing exposure to cues for the behaviour (12.3)	
		Distraction (12.4)	
		Body changes (12.6)	
	Comparison of behaviour	Demonstration of the behaviour (6.1)	
	Feedback and monitoring	Monitoring of behaviour by others without feedback (2.1)	
		Monitoring of outcomes of behaviour without feedback (2.5)	
		Biofeedback (2.6)	
	Identity	Incompatible beliefs (13.3)	Identification of self as role model (13.1)
		Valued self-identity (13.4)	
		Identity associated with changed behaviour (13.5)	
Prominent determinants 'knowledge', 'memory', 'social influences' and 'beliefs about consequences'	Covert learning	Imaginary punishment (16.1)	
		Imaginary reward (16.2)	
	Comparison of outcomes	Comparative imagining of future outcomes (9.3)	
	Natural consequences	Monitoring of emotional consequences (5.4)	Anticipated regret (5.5)
		Information about emotional consequences (5.6)	
	Shaping knowledge	Behavioural experiments (4.4)	
	Goals and planning		Discrepancy between current behaviour and goal (1.6)
	Repetition and substitution	Behavioural practice/rehearsal (8.1)	
		Behaviour substitution (8.2)	
		Habit reversal (8.4)	
		Overcorrection (8.5)	
		Generalisation of target behaviour (8.6)	
	Associations	Cue signalling reward (7.2)	
		Reduce prompts/cues (7.3)	
		Remove access to the reward (7.4)	
		Remove aversive stimulus (7.5)	

TABLE 18 Behaviour change techniques excluded from intervention development or intended but not subsequently operationalised (WP3b) (continued)

BCTs for changing determinants of behaviour ⁵²		BCTs excluded because of delivery mechanism or contextual constraints (BCT taxonomy code reference) ⁶⁶	BCTs intended but not operationalised (BCT taxonomy code reference) ⁶⁶
	Regulation	Satiation (7.6)	
		Exposure (7.7)	
		Associative learning (7.8)	
		Pharmacological support (11.1)	
		Reduce negative emotions (11.2)	
		Paradoxical instructions (11.4)	
	Reward and threat	Material incentive (behaviour) (10.1)	
		Material reward (behaviour) (10.2)	
		Non-specific reward (10.3)	
		Social incentive (10.5)	
		Non-specific incentive (10.6)	
		Self-incentive (10.7)	
		Incentive (outcome) (10.8)	
		Self-reward (10.9)	
		Reward (outcome) (10.10)	
		Future punishment (10.11)	
Less evident determinants 'self-belief' and 'scheduled consequences'	Self-belief	Mental rehearsal of successful performance (15.2)	Verbal persuasion about capability (15.1)
		Self-talk (15.4)	
	Scheduled consequences	Behavioural cost (14.1)	
		Punishment (14.2)	
		Remove reward (14.3)	
		Reward approximation (14.4)	
		Rewarding completion (14.5)	
		Situation-specific reward (14.6)	
		Reward incompatible behaviour (14.7)	
		Reward alternative behaviour (14.8)	
		Reduce reward frequency (14.9)	
		Remove punishment (14.10)	

Reproduced from Glidewell *et al.*⁷¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

The 2014–15 QOF measured achievement against 81 indicators; practices scored points on the basis of achievement against each indicator, up to a maximum of 559.

All patients registered at ASPIRE practices at baseline are included in *Table 21*.

TABLE 19 Cluster-randomised evaluation: characteristics of the non-randomised and randomised practices (WP4a)

Characteristic	Non-randomised practices (n = 100)	Randomised practices (n = 178)
Practice list size		
Mean (SD)	7217.35 (3369.46)	7249.76 (4306.72)
Median (range)	6961.50 (1622.00–15,290.00)	6565.50 (1268.00–25,495.00)
Missing	6	0
Number of GP partners (headcount)^a		
Mean (SD)	5.13 (3.07)	4.06 (3.07)
Median (range)	5.00 (1.00–13.00)	4.00 (0.00–15.00)
Missing	21	29
Number of GP partners (FTE)^a		
Mean (SD)	4.17 (2.45)	3.27 (2.49)
Median (range)	3.67 (0.44–10.23)	2.56 (0.00–11.41)
Missing	21	29
Number of salaried GPs (headcount)^b		
Mean (SD)	0.58 (1.14)	0.93 (1.44)
Median (range)	0.00 (0.00–7.00)	0.00 (0.00–6.00)
Missing	21	29
Number of salaried GPs (FTE)^b		
Mean (SD)	0.36 (0.73)	0.65 (1.04)
Median (range)	0.00 (0.00–4.27)	0.00 (0.00–4.37)
Missing	21	29
Number of GPs (headcount)^b		
Mean (SD)	5.71 (3.18)	4.99 (3.37)
Median (range)	5.00 (1.00–13.00)	4.00 (1.00–15.00)
Missing	21	29
Number of GPs (FTE)^b		
Mean (SD)	4.52 (2.49)	3.92 (2.72)
Median (range)	4.27 (0.44–10.23)	3.22 (0.61–11.41)
Missing	21	29

TABLE 19 Cluster-randomised evaluation: characteristics of the non-randomised and randomised practices (WP4a) (continued)

Characteristic	Non-randomised practices (n = 100)	Randomised practices (n = 178)
Deprivation score (IMD 2015)^c		
Mean (SD)	31.32 (12.23)	30.85 (13.70)
Median (range)	31.49 (6.42–59.74)	30.22 (5.54–57.72)
Missing	9	3
% patients who would recommend practice		
Mean (SD)	76.06 (14.46)	74.33 (14.11)
Median (range)	79.29 (16.98–96.70)	76.59 (34.35–98.46)
Missing	9	4
% patients who saw/spoke to nurse or GP same or next day^d		
Mean (SD)	52.34 (12.23)	51.94 (15.61)
Median (range)	52.57 (24.28–77.27)	52.93 (13.13–89.72)
Missing	9	3
QOF score^e		
Mean (SD)	532.63 (29.16)	531.11 (31.70)
Median (range)	541.42 (384.54–559.00)	539.61 (336.07–559.00)
Missing	7	1
Teaching practice?^f		
Yes, n (%)	33 (33.0)	67 (37.6)
No, n (%)	67 (67.0)	111 (62.4)

FTE, full-time equivalent.

a The number of GP partners (headcount or FTE) has been derived as the difference between the total number of GPs and the number of salaried GPs in each practice.

b GP staff data have been taken from the NHS Digital General and Personal Medical Services data set⁸⁵ and reflect the general practice workforce at 30 September 2015.

c The deprivation scores are a population-weighted average over the IMD 2015 scores of the Lower Layer Super Output Areas in which the practice population lives. These scores take a positive value where a higher score indicates a more deprived area. The deprivation data have been obtained from the Public Health England *National General Practice Profiles*⁴⁵ for the year 2015.

d Patient satisfaction scores have been obtained from the Public Health England *National General Practice Profiles*⁴⁵ for the financial year 2014/15.

e The 2014/15 QOF score measured achievement against 81 indicators; practices scored points on the basis of achievement against each indicator, up to a maximum of 559.

f All known teaching practices in West Yorkshire (as of 10 January 2017) are categorised as 'Yes', otherwise they have been set to 'No'.

TABLE 20 Cluster-randomised evaluation: general practice characteristics at baseline by trial arm (WP4a)

	Trial 1		Trial 2			
	Diabetes control (n = 40)	Risky prescribing (n = 40)	BP control (n = 32)	Anticoagulation in AF (n = 32)	Non-intervention (n = 34)	Total (n = 178)
List size						
Mean (SD)	7084.35 (3786.47)	7175.75 (3857.00)	7538.94 (4932.93)	7421.28 (4171.20)	7097.82 (5057.86)	7249.76 (4306.72)
Median (range)	6703.00 (14,33.00–14,822.00)	6764.50 (14,93.00–14,760.00)	6367.50 (1268.00–17,429.00)	7067.00 (1889.00–18,891.00)	6074.50 (1723.00–25,495.00)	6565.50 (1268.00–25,495.00)
Overall QOF score^a						
Mean (SD)	535.40 (29.93)	531.06 (35.79)	527.17 (27.03)	532.99 (21.75)	527.92 (40.34)	531.11 (31.70)
Median (range)	542.21 (387.44–559.00)	542.17 (389.19–559.00)	534.66 (447.18–557.33)	538.89 (461.17–559.00)	536.58 (336.07–559.00)	539.61 (336.07–559.00)
Pre-intervention achievement						
<i>Diabetes control</i>						
Mean (SD) (%)	32.87 (6.94)	34.33 (7.72)	32.50 (7.10)	33.37 (5.48)	33.24 (7.96)	33.29 (7.08)
<i>Risky prescribing</i>						
Mean (SD) (%)	7.92 (5.14)	7.92 (3.61)	7.27 (3.64)	7.92 (2.53)	8.47 (4.82)	7.91 (4.07)
<i>BP control</i>						
Mean (SD) (%)	66.53 (6.42)	66.42 (7.02)	65.91 (7.53)	65.27 (6.25)	64.96 (6.32)	65.87 (6.67)
<i>Anticoagulation in AF</i>						
Mean (SD) (%)	66.50 (14.42)	67.30 (8.43)	66.54 (10.80)	66.35 (8.29)	64.08 (15.48)	66.20 (11.82)
a There was one practice with a missing value for overall QOF score in trial 2.						

TABLE 21 Cluster-randomised evaluation: patient characteristics at baseline by trial arm (all patients) (WP4a)

	Trial 1		Trial 2		Non-intervention (n = 243,856)	Total (n = 1,311,258)
	Diabetes control (n = 288,130)	Risky prescribing (n = 290,407)	BP control (n = 249,571)	Anticoagulation in AF (n = 239,294)		
Age						
Mean (SD)	38.03 (22.88)	37.60 (23.14)	39.36 (23.20)	38.99 (23.22)	36.93 (22.48)	38.16 (23.00)
Median (range)	36.00 (0.00–107.00)	36.00 (0.00–106.00)	39.00 (0.00–105.00)	39.00 (0.00–107.00)	35.00 (0.00–108.00)	37.00 (0.00–108.00)
Gender						
Female	141,328 (49.05%)	144,426 (49.73%)	124,824 (50.02%)	120,289 (50.27%)	120,680 (49.49%)	651,547 (49.69%)
Male	146,799 (50.95%)	145,980 (50.27%)	124,746 (49.98%)	119,003 (49.73%)	123,169 (50.51%)	659,697 (50.31%)
Indeterminate	2 (0.00%)	1 (0.00%)	1 (0.00%)	1 (0.00%)	7 (0.00%)	12 (0.00%)
Unknown	1 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.00%)	0 (0.00%)	2 (0.00%)
Comorbidity						
0–3	276,280 (95.89%)	277,184 (95.45%)	239,455 (95.95%)	229,329 (95.84%)	234,116 (96.01%)	1,256,364 (95.81%)
4 +	11,850 (4.11%)	13,223 (4.55%)	10,116 (4.05%)	9965 (4.16%)	9740 (3.99%)	54,894 (4.19%)

Adapted from Willis *et al.*⁸⁷ © 2020 Willis *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

TABLE 22 Primary outcome analysis: diabetes control (WP4a)

Variable	Comparison	Estimate	SE	OR	Lower limit of 97.5% CI for OR	Upper limit of 97.5% CI for OR	F-value	Degrees of freedom	p-value
Allocation	Diabetes control implementation package vs. control (risky prescribing package)	0.0251	0.0635	1.0254	0.8894	1.1822	0.16	1	0.6926
Gender	Female vs. male	-0.1237	0.0270	0.8837	0.8317	0.9389	20.91	1	<0.0001
Age (years)		0.0219	0.0010	1.0221	1.0198	1.0245	454.34	1	<0.0001
Relative to 100 patient increase in list size at baseline		0.0006	0.0010	1.0006	0.9984	1.0028	0.38	1	0.5400
CCG	NHS Airedale, Wharfedale and Craven CCG vs. NHS Wakefield CCG	-0.1485	0.1712	0.8620	0.5873	1.2652	1.65	9	0.0940
CCG	NHS Bradford City CCG vs. NHS Wakefield CCG	0.2278	0.1607	1.2559	0.8761	1.8003			
CCG	NHS Bradford Districts CCG vs. NHS Wakefield CCG	0.1804	0.1130	1.1976	0.9297	1.5428			
CCG	NHS Calderdale CCG vs. NHS Wakefield CCG	0.1343	0.1570	1.1437	0.8044	1.6261			
CCG	NHS Greater Huddersfield CCG vs. NHS Wakefield CCG	0.1470	0.1581	1.1584	0.8128	1.6510			
CCG	NHS Leeds North CCG vs. NHS Wakefield CCG	0.4710	0.1628	1.6016	1.1120	2.3067			
CCG	NHS Leeds South and East CCG vs. NHS Wakefield CCG	0.1329	0.1362	1.1421	0.8417	1.5497			
CCG	NHS Leeds West CCG vs. NHS Wakefield CCG	0.3552	0.1557	1.4265	1.0062	2.0224			
CCG	NHS North Kirklees CCG vs. NHS Wakefield CCG	0.2226	0.1415	1.2494	0.9098	1.7157			
Pre-intervention achievement for risky prescribing		0.0054	0.0082	1.0054	0.9870	1.0241	0.43	1	0.5123
Pre-intervention achievement for diabetes control		0.0305	0.0048	1.0309	1.0199	1.0420	40.50	1	<0.0001
Overall QOF score		-0.0016	0.0013	0.9984	0.9954	1.0013	1.52	1	0.2174
Comorbidity		-0.0378	0.0410	0.9629	0.8783	1.0557	0.85	1	0.3572

TABLE 23 Primary outcome analysis: risky prescribing (WP4a)

Variable	Comparison	Estimate	SE	OR	Lower limit of 97.5% CI for OR	Upper limit of 97.5% CI for OR	F-value	Degrees of freedom	p-value
Allocation	Risky prescribing implementation package vs. control (diabetes control package)	-0.2046	0.0855	0.8150	0.6729	0.9871	5.73	1	0.0167
Gender	Female vs. male	-0.1031	0.0451	0.9021	0.8153	0.9981	5.22	1	0.0224
Age		-0.0125	0.0017	0.9876	0.9838	0.9914	53.35	1	<0.0001
Relative to 100 patient increase in list size at baseline		-0.0001	0.0013	0.9999	0.9969	1.0029	0.01	1	0.9258
CCG	NHS Airedale, Wharfedale and Craven CCG vs. NHS Wakefield CCG	0.1944	0.2141	1.2146	0.7517	1.9626	1.72	9	0.0792
CCG	NHS Bradford City CCG vs. NHS Wakefield CCG	0.1304	0.2245	1.1392	0.6888	1.8841			
CCG	NHS Bradford Districts CCG vs. NHS Wakefield CCG	0.0901	0.1460	1.0943	0.7888	1.5179			
CCG	NHS Calderdale CCG vs. NHS Wakefield CCG	0.0940	0.1987	1.0985	0.7036	1.7151			
CCG	NHS Greater Huddersfield CCG vs. NHS Wakefield CCG	0.3294	0.2027	1.3901	0.8826	2.1894			
CCG	NHS Leeds North CCG vs. NHS Wakefield CCG	-0.1277	0.2180	0.8801	0.5400	1.4346			
CCG	NHS Leeds South and East CCG vs. NHS Wakefield CCG	-0.2525	0.1820	0.7769	0.5167	1.1681			
CCG	NHS Leeds West CCG vs. NHS Wakefield CCG	-0.2685	0.2134	0.7645	0.4739	1.2335			
continued									

TABLE 23 Primary outcome analysis: risky prescribing (WP4a) (continued)

Variable	Comparison	Estimate	SE	OR	Lower limit of 97.5% CI for OR	Upper limit of 97.5% CI for OR	F-value	Degrees of freedom	p-value
CCG	NHS North Kirklees CCG vs. NHS Wakefield CCG	0.0699	0.1823	1.0724	0.7127	1.6137			
Pre-intervention achievement for risky prescribing		0.0778	0.0109	1.0809	1.0549	1.1076	51.13	1	<0.0001
Pre-intervention achievement for diabetes control		-0.0058	0.0066	0.9942	0.9797	1.0090	0.77	1	0.3792
Overall QOF score		0.0002	0.0018	1.0002	0.9961	1.0043	0.01	1	0.9040
Comorbidity		0.0294	0.0575	1.0298	0.9052	1.1716	0.26	1	0.6094

TABLE 24 Primary outcome analysis: BP control (WP4a)

Variable	Comparison	Estimate	SE	OR	Lower limit of 97.5% CI for OR	Upper limit of 97.5% CI for OR	F-value	Degrees of freedom	p-value
Allocation	BP control package vs. control (anticoagulation in AF package)	0.0520	0.0419	1.0533	0.9589	1.1571	1.54	1	0.2151
Gender	Female vs. male	0.1262	0.0136	1.1345	1.1005	1.1696	86.39	1	<0.0001
Age		0.0265	0.0006	1.0268	1.0256	1.0281	2304.94	1	<0.0001
Relative to 100 patient increase in list size at baseline		-0.0014	0.0006	0.9986	0.9974	0.9999	6.19	1	0.0128
CCG	NHS Airedale, Wharfedale and Craven CCG vs. NHS Wakefield CCG	0.0754	0.0958	1.0783	0.8700	1.3367	2.43	8	0.0126
CCG	NHS Bradford City CCG vs. NHS Wakefield CCG	-0.1343	0.0848	0.8743	0.7229	1.0574			
CCG	NHS Calderdale CCG vs. NHS Wakefield CCG	0.2631	0.0936	1.3010	1.0549	1.6045			
CCG	NHS Greater Huddersfield CCG vs. NHS Wakefield CCG	0.0079	0.0874	1.0079	0.8286	1.2259			
CCG	NHS Leeds North CCG vs. NHS Wakefield CCG	0.1930	0.1035	1.2129	0.9619	1.5295			
CCG	NHS Leeds South and East CCG vs. NHS Wakefield CCG	0.0518	0.0811	1.0532	0.8782	1.2631			
CCG	NHS Leeds West CCG vs. NHS Wakefield CCG	0.0423	0.0804	1.0432	0.8711	1.2493			
CCG	NHS North Kirklees CCG vs. NHS Wakefield CCG	0.1343	0.0827	1.1437	0.9501	1.3768			
Pre-intervention achievement for anticoagulation in AF		-0.0039	0.0025	0.9961	0.9905	1.0018	2.33	1	0.1266
Pre-intervention achievement for BP control		0.0262	0.0036	1.0265	1.0184	1.0348	54.11	1	<0.0001
Overall QOF score		0.0007	0.0010	1.0007	0.9986	1.0029	0.59	1	0.4425
Comorbidity		-0.0008	0.0227	0.9992	0.9496	1.0513	0.00	1	0.9713

TABLE 25 Primary outcome analysis: anticoagulation in AF (WP4a)

Variable	Comparison	Estimate	SE	OR	Lower limit of 97.5% CI for OR	Upper limit of 97.5% CI for OR	F-value	Degrees of freedom	p-value
Allocation	Anticoagulation in AF package vs. control (BP control package)	-0.1035	0.0833	0.9017	0.7482	1.0868	1.54	1	0.2141
Gender	Female vs. male	-0.0618	0.0566	0.9401	0.8280	1.0673	1.19	1	0.2752
Age		-0.0117	0.0029	0.9883	0.9819	0.9948	16.35	1	<0.0001
Relative to 100 patient increase in list size at baseline		-0.0005	0.0012	0.9995	0.9969	1.0022	0.16	1	0.6907
CCG	NHS Airedale, Wharfedale and Craven CCG vs. NHS Wakefield CCG	-0.0559	0.1582	0.9456	0.6632	1.3483	0.72	8	0.6724
CCG	NHS Bradford City CCG vs. NHS Wakefield CCG	-0.4125	0.2221	0.6620	0.4024	1.0891			
CCG	NHS Calderdale CCG vs. NHS Wakefield CCG	0.0478	0.1765	1.0489	0.7061	1.5581			
CCG	NHS Greater Huddersfield CCG vs. NHS Wakefield CCG	-0.0815	0.1758	0.9217	0.6215	1.3670			
CCG	NHS Leeds North CCG vs. NHS Wakefield CCG	0.1597	0.1762	1.1732	0.7904	1.7414			
CCG	NHS Leeds South and East CCG vs. NHS Wakefield CCG	-0.0243	0.1549	0.9760	0.6896	1.3813			
CCG	NHS Leeds West CCG vs. NHS Wakefield CCG	0.0334	0.1501	1.0340	0.7386	1.4476			
CCG	NHS North Kirklees CCG vs. NHS Wakefield CCG	0.0427	0.1528	1.0436	0.7409	1.4699			
Pre-intervention achievement for anticoagulation in AF		0.0372	0.0056	1.0379	1.0248	1.0511	43.52	1	<0.0001
Pre-intervention achievement for BP control		0.0013	0.0071	1.0013	0.9855	1.0173	0.03	1	0.8593
Overall QOF score		0.0026	0.0021	1.0026	0.9978	1.0074	1.46	1	0.2276
Comorbidity		0.0218	0.0544	1.0220	0.9047	1.1545	0.16	1	0.6887

TABLE 26 Number of practices in each trial arm where at least one outreach visit took place (WP5)

At least one outreach visit held	Diabetes control (n = 40)	Risky prescribing (n = 40)	BP control (n = 32)	Anticoagulation in AF (n = 32)	Total (n = 144)
Yes	20 (50.00%)	25 (62.50%)	11 (34.38%)	11 (34.38%)	67 (46.53%)
No	20 (50.00%)	15 (37.50%)	21 (65.63%)	21 (65.63%)	77 (53.47%)

TABLE 27 Number of outreach visits held at each practice (by trial arm) (WP5)

Number of outreach visits held	Diabetes control (n = 40)	Risky prescribing (n = 40)	BP control (n = 32)	Anticoagulation in AF (n = 32)	Total (n = 144)
0	20 (50.00%)	15 (37.50%)	21 (65.63%)	21 (65.63%)	77 (53.47%)
1	18 (45.00%)	22 (55.00%)	10 (31.25%)	9 (28.13%)	59 (40.97%)
2	2 (5.00%)	3 (7.50%)	1 (3.13%)	2 (6.25%)	8 (5.56%)

TABLE 28 Intervention fidelity level of practices by trial arm (WP4a)

Intervention fidelity level ^a	Diabetes control (n = 40)	Risky prescribing (n = 40)	BP control (n = 32)	Anticoagulation in AF (n = 32)	Total (n = 144)
High	18 (45.00%)	8 (20.00%)	11 (34.38%)	10 (31.25%)	47 (32.64%)
Medium	22 (55.00%)	32 (80.00%)	12 (37.50%)	16 (50.00%)	82 (56.94%)
Low	0 (0.00%)	0 (0.00%)	9 (28.13%)	6 (18.75%)	15 (10.42%)

a Intervention fidelity levels are based on the following criteria:

1. A&F reports received at each time point
2. outreach visit 1 undertaken
3. joined organisational group for searches
4. joined group for prompts (risky prescribing only).

Where the levels are defined as:

- High – all of 1–3 achieved (and 4 for risky prescribing practices).
- Medium – achieved 1 and at least one other from 2–4.
- Low – achieved 1 only.

TABLE 29 Summary of delivery and receipt of intervention (WP4a)

Intervention components		Delivery	Receipt
A&F	Reports	Reports were sent by e-mail and post every 3 months, and taken along to outreach visits	All trial practices received reports as intended (with the exception of those that closed/merged) All of the process evaluation practices received the reports
	Searches	Practices were sent an e-mail inviting them to join an organisational group through SystmOne	126 (87.5%) trial practices joined the organisational group 75.0% of responses from trial practices stated that they downloaded and used the searches when asked in the post-trial survey

continued

TABLE 29 Summary of delivery and receipt of intervention (WP4a) (continued)

Intervention components		Delivery	Receipt
Educational outreach	Significant event audit forms	Forms were sent by post and taken along to educational outreach visits	<p>For the process evaluation practices, 75.0% of practices joined the organisational group and all of those that responded to the post-trial survey indicated that they had downloaded the searches</p> <p>Receipt was not tracked in the trial practices</p> <p>In the relevant process evaluation practices ($n = 4$), there was evidence of receipt in one practice but no evidence of receipt in the others</p>
	Initial educational outreach meeting	Practices were offered an educational outreach visit on every report. Every practice was also contacted by an administrator by telephone and e-mail to arrange a date	<p>Sixty-seven (46.53%) trial practices took up the offer of an outreach visit. Reasons for declining the offer included not interested, too busy, do not need it, permanent closure of practice</p> <p>Within the process evaluation, 87.5% of practices took up the offer; the practice that declined said that it felt confident about what it needed to do and was able to do it itself</p> <p><i>I think we just felt that we were pretty much on the ball with it (. . .) we didn't need a pharmacist to push us. I think we decided that we already had enough resources to get on with it . . .</i></p> <p><i>Valley practice (AF), interview with GP partner, lead for cardiovascular disease</i></p>
	Outreach support	<p>Practices were offered outreach support as part of the initial outreach meeting. Outreach support was offered to all practices. The possibility of support was also mentioned on each report</p> <p>Outreach support was either delivered face to face or remotely via e-mail</p>	<p>Outreach support was taken up by 77 (53.47%) trial practices</p> <p>Most support was provided remotely by a pharmacist external to the practice and took the form of (a) running searches for agreed indicators, (b) reviewing patient notes for a number of relevant indicators and (c) creating a list of recommendations for management</p> <p>In the process evaluation practices, two practices had face-to-face outreach support, three practices received support remotely, two declined support and the final practice was not offered support as a result of declining the educational outreach visit</p>
	Second outreach meeting	<p>All practices that took up an outreach visit were offered a second outreach meeting by an administrator</p> <p>Five out of eight process evaluation practices were offered a second visit; resources and time constraints meant that not all practices were contacted by the end of the trial</p>	<p>Eight (5.56%) trial practices had a second outreach visit</p> <p>Three process evaluation practices had a second outreach meeting. Out of the remaining two that were offered a visit, one practice declined it as it felt it did not need it and one practice said yes but then was unable to set a date for the meeting before the trial ended</p>

TABLE 29 Summary of delivery and receipt of intervention (WP4a) (continued)

Intervention components		Delivery	Receipt
Reminders	Computerised protocols	Practices were sent an e-mail inviting them to join an organisational group through SystmOne	Eight (32.00%) trial practices joined the organisational group that gave access to the risky prescribing computerised protocol For risky prescribing, the protocols were downloaded by both of the process evaluation practices
	Physical reminders	Pens and sticky notes were sent to all practices by post (with the second A&F form) in August 2015	Receipt was not tracked in the trial practices
		Laminates for BP were sent to the relevant practices by post (with the second A&F form) in August 2015 Laminates for anticoagulation in AF were sent to the relevant practices by post in January 2016	All eight process evaluation practices received the pens and sticky notes. None of the process evaluation practices remembered receiving a laminate for BP ($n = 2$) or anticoagulation in AF ($n = 2$) and the laminates were not seen by CH while visiting the practices

TABLE 30 Perceptions of implementation package components (practices anonymised and assigned pseudonyms) (WP5)

Intervention component		Content	
A&F	Reports	Persuasive/credible	<i>[GP partner] commented on the quote on the side, mentioning that [the expert quoted] was a known atrial fibrillation expert – knows what he is talking about ...</i> <i>Observation of meeting, Valley</i>
		Reinforcement/reminder	<i>I think the fact that the emphasis has come, the challenge has come externally has been quite useful for me because it gets quite boring seen as being the negative person that always says that we should do better (...) so it's been good in terms of getting people thinking, the conversation going ...</i> <i>GP partner/diabetes lead interview, Lake</i>
		Informational (anticoagulation arm only)	<i>It's just having the ASPIRE presence, it's just a reminder, it's just made us think, it's just brought it to the forefront of my mind, because what we know what to do, it's not changed that, what we should do, it's just doing it</i> <i>Final practice meeting, Brook</i> <i>There was a bit of exclamation around the five times higher risk of stroke part. As they read through the first page, [GP clinical lead] said that it was good they were reading through the whole thing 'so that we're all up-skilled'...</i> <i>Observation of meeting, Valley</i>

continued

TABLE 30 Perceptions of implementation package components (practices anonymised and assigned pseudonyms) (WP5) (continued)

Intervention component		Content	
Educational Outreach	Searches	Motivational (stimulated competitiveness)	<p>A GP said we are not that bad on the graph, but it's not where the rest of our therapies are, which annoys me, 'people are above us – I don't like it!' Another GP agreed, we are competitive ...</p> <p><i>Educational outreach, Lake</i></p>
		Enabling (resource-saving)	<p>It was a big help, it saves us having to put all the searches together ...</p> <p><i>Interview with practice manager, Brook</i></p> <p>P: I have run the searches. I: (...) have they been useful? P: Yeah, yeah 'cos I wouldn't have known what to put [in them] ...</p> <p><i>Interview with practice manager, Treetop</i></p>
	Initial educational outreach meeting	Enabling (gateway to further support)	<p>I asked whether they had had any time to look at the audit and feedback form and she said no, they didn't have time, they had been waiting for the support to come ...</p> <p><i>Conversation with practice manager, Flower</i></p> <p>They decided they would take the outreach session up to help them get going ...</p> <p><i>Meeting, Dale</i></p>
		Motivational (focused attention on topic)	<p>It was 'nice to have the opportunity to review' what they were doing and 'good to have a pharmacist on board' because they 'don't get input from them often'. It was 'good to have an extra check on how we prescribe'</p> <p><i>Telephone call with GP partner/clinical lead, Brook</i></p> <p>The outreach meeting was 'really useful' as it is always more helpful to have someone come in and be there to ask questions of [them] and 'make what needs to be done clear' (...) It 'needs to be on an agenda to bring it up'. This was part of the reason that being involved with ASPIRE was good, because it put AF on the agenda and helped keep it on the agenda long enough for the work to get done'</p> <p><i>Field notes from interview, practice manager, Flower</i></p>
		Promoting whole practice engagement	<p>I: What difference does [having a group meeting] make? P: a consistent approach (...) less chance of people falling through the net you know in error by mistake, here in a hurry cos somebody else is always looking at the prescriptions ...</p> <p><i>Interview with GP partner/clinical lead, Treetop</i></p>
		Clarifying actions	<p>The chair said 'yes, I think it made it much clearer what the risk is, that you were actually saving people's lives by anticoagulating' ...</p> <p><i>Final practice meeting, Flower</i></p>

TABLE 30 Perceptions of implementation package components (practices anonymised and assigned pseudonyms) (WP5) (continued)

Intervention component		Content	
Outreach support	Enabling (adding capacity)		<i>[Educational outreach] made things a lot clearer (...) just going through the categories and knowing what to look for...</i> <i>GP partner/clinical lead, Treetop</i>
			<i>The key thing for her was the additional support – it was a huge help. It's much easier for them to look at the sheet and get on with the work now that the support had done the lists, than it is to wait for a GP to have the time to run the searches...</i> <i>Field notes, interview with practice manager, Brook</i>
	Not needed (using own resources)		<i>GP lead said we should do what we can first, and then take up the support as needed. He said, the ones we can't do, we will ask him to look at...</i> <i>Educational outreach, Treetop</i>
Reminders	Computerised protocols (risky prescribing only)	Enabling (makes remembering to act easier)	<i>I'm more aware (...) It stays there. The medication stays in my head and it prompts me when I see that medication (...) Needs protection or need to see a doctor to talk about this...</i> <i>Interview with prescribing clerk, Treetop</i>
		Not needed (too many protocols)	<i>The PM [practice manager] said she thought they'd decided to not use them because there are so many of those types of things on there, people get a bit fed up of seeing...</i> <i>Final practice meeting, Brook</i>

TABLE 31 Reported actions in response to implementation package components (WP5)

Intervention component		Response	
A&F	Reports	In response to initial feedback	<i>I think the initial body of feedback where we were right down the bottom (...) my practice really doesn't like that. They are quite competitive in a lot of ways (...) to be right down at the bottom wasn't acceptable so we had to do something about it. That was just the sort of general (...) feeling'</i> <i>Interview with GP, Valley</i>
		Social comparison – desire to do better	<i>We realised where we were deficient and how we compare to other people, it's good to have other people to compare against 'cos you're actually working quite a lot in isolation in a way as a practice...</i> <i>Interview with GP, Treetop</i>
		Review current practice processes	<i>They had decided to use this [peer review meeting] as a means to re-think their approach to hypertension as a result of ASPIRE [reports]...</i> <i>Peer review meeting, Hill</i>

continued

TABLE 31 Reported actions in response to implementation package components (WP5) (continued)

Intervention component	Response
	<p>Positive responses to repeat feedback</p> <p>Discussing progress at meetings</p> <p>Becoming more confident prescribing</p> <p>Negative response to repeat feedback</p> <p>Reinterpretation of data</p> <p>Limiting exposure to feedback</p> <p>Demotivation</p> <p>Searches</p> <p>Adapted searches to practice needs</p> <p>Repeated use (stimulated by reminders)</p> <p>Limited use (lack of capacity)</p>
	<p><i>The GPs said they discussed the reports, it possibly raised the consciousness, but that's it ...</i></p> <p><i>Final meeting, River</i></p> <p><i>I think when this started, as the male GP was saying, we would have referred everybody whereas I think I now feel more comfortable starting people [on anticoagulants] ...</i></p> <p><i>Final practice meeting, Flower</i></p> <p><i>He said that they had felt like they had done quite a lot of work but this was not reflected in the figures. He laughed as he said it, it felt a bit dispiriting really. He felt they were doing so much work just to stay in the same place. Other people nodded and agreed ...</i></p> <p><i>Final practice meeting, River</i></p> <p><i>I mean to be honest with you normally we'd share it with all the partners, but because the results didn't look that good to me, I didn't want to embarrass [GP – diabetes lead] by giving it to all the partners ...</i></p> <p><i>Practice manager interview, Dale</i></p> <p><i>Erm I think the reports were interesting erm and I think it becomes a little bit demoralising when you see that things are just getting worse ...</i></p> <p><i>GP lead interview, Lake</i></p> <p><i>He explained that the GPs could not look at the searches as they were, as they were too long (...) targeting too many patients, they didn't have the resources (...) [they] refined the searches (...) [so that] about 30 people were on the list given to the diabetes lead ...</i></p> <p><i>Final practice meeting, River</i></p> <p><i>We did, we searched once every month (...) And then we reviewed, brought in all those patients in that we hadn't ... treated that were on the recall list that we hadn't treated, and we reviewed them. So, although it's more work (...) We were on top of it ...</i></p> <p><i>Second interview, GP lead, April 2016</i></p> <p><i>The GP lead said she did look at them, she used the renal one, but that was the only one she used. The medical student used them at the beginning a lot to have a look at what was going on ...</i></p> <p><i>Final practice meeting, Hill</i></p> <p><i>[GP partner said] we have got good systems, patients do get reviewed (...) we need to make sure [BP] doesn't break our systems ...</i></p> <p><i>Educational outreach, Lake</i></p>

TABLE 31 Reported actions in response to implementation package components (WP5) (continued)

Intervention component		Response
Educational outreach		Limited use (lack of awareness)
		None of them had seen any searches from ASPIRE either, although they had done their own sift through the AF QOF list last time ... <i>Meeting with GP lead, Valley</i>
		Limited use (queried value of work)
		... there was a discussion about we've got these searches to do but is it worth doing, I think there was a valid judgement about what we were going to get out from doing it ... <i>Interview with practice manager, Lake</i>
	Initial educational outreach meeting	Agreeing an action plan and assigning tasks
		It looks like about 45 people (...) that doesn't sound too onerous does it? [Staff] were nodding at this ... [clinical lead] could do the list review (...) should not be too difficult ... <i>Educational outreach, Flower</i>
		Waiting for external support to carry out action plan (ASPIRE or other)
		It worked really well while I had my student over the summer ... I think we made a massive improvement at the beginning and then it's sort of tapered off as [we] just couldn't keep the momentum going I think ... <i>Interview with GP partner/clinical lead, Hill</i>
		Protecting limited resources (resisting action-planning)
		... we've kind of waited for [outreach support] to happen (...) I hadn't appreciated that we actually needed to be chasing that up and organising it! <i>Interview with GP partner/clinical lead, Dale</i>
		Protecting limited resources (resisting action-planning)
		At the outreach meeting, the practice had discussed adding in some hypertension work during the flu work and he said yes, he remembered, but that was wildly unrealistic <i>Field notes from interview with GP partner, Lake</i>
Outreach support	Rationalising workload from support	Doctors generally – unless well-resourced clinically – won't take this on – someone else down the line has to take it on (...) a list, like our searches generated (...) doesn't tell you what to focus on, how do you decide? Someone has to think about it and interpret it <i>Feedback from practice manager, River</i>
		[It] was targeting too many patients, they didn't have the resources. The chair agreed, when you see a list to review of about 100 patients, your heart sinks. The PM [practice manager] said we refined the searches, then the pharmacist looked at it, and then about 30 people were on the list given to the diabetes lead so he could look if it was clinically worthwhile to doing anything with them. He said that they just don't have the time or capacity ... <i>Final meeting, River</i>

continued

TABLE 31 Reported actions in response to implementation package components (WP5) (continued)

Intervention component		Response	
		Acting on recommendations	<p>He said that the input from the pharmacist was good, there was only one error in about 100 records (...) He said that most of the comments by the pharmacist were relevant and he acted on most of the suggestions for change ...</p> <p><i>Conversation with GP partner/clinical lead, Brook</i></p>
		Dissatisfaction/disengagement	<p>The GP said (...) if the impression is given by an outside agency of right, OK, that there's a project we're keen for you to be involved in, we put up our hands and say we'll be involved, and then they come and say how do you need support, and we say these are the things that could help, and they say right we'll deliver on those things, I think the tendency then – and I don't know about other practices – is to await that action and then we will move it along and that bit is where it fell down ...</p> <p><i>Final practice meeting, Dale</i></p> <p>... there was a discussion about something being then highlighted onto people's records (...) and I don't think that ever appeared (...) something that would automatically appear on people's records ...</p> <p><i>Interview with GP partner/clinical lead, Hill</i></p>
	Reminders	Computerised protocols (risky prescribing only)	<p>Used as additional check on prescribing practice</p> <p>He felt the top few indicators were something he did routinely, the message had definitely stuck there, but the bottom few indicators were not routine. He felt that for those, the [protocols] were still something useful to have</p> <p><i>Interview with GP, Treetop</i></p>
		Ignored or deleted	<p>I don't like all the flashing lights, I don't like the prompts and the templates, I think it's so complicated that it makes your job, it's ridiculous, you wouldn't have a pilot looking at something like this would you, they wouldn't be able to concentrate ...</p> <p><i>Interview with GP, Treetop</i></p>

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library